

AIDS and the private sector

It will take more than awareness to slow the HIV pandemic in Africa.

This week, the rock star Bono and a host of other celebrities are promoting the US launch of 'Product Red', the private sector's campaign to fight AIDS. Some of the profits on products sold in this initiative, as agreed by their manufacturers, will go to the Global Fund to Fight AIDS, Tuberculosis and Malaria, specifically to confront AIDS in Africa. The arrival of iPods, phones and clothing that will highlight the AIDS pandemic and provide financial support for the global fund is to be welcomed.

The publicity surrounding the US launch — which follows a similar process earlier this year in Britain — may convey the impression that large global corporations are at the forefront of the fight against AIDS. But a closer look at the situation reveals that this is rarely the case (see page 738).

For two decades, AIDS activists and officials have implored the private sector to join the fight against AIDS. In reality, that effort remains overwhelmingly dependent on public funding and public officials. In all but a few cases, the private sector has been a minor player. The Product Red campaign is one sign that this may finally be changing — but much more needs to be done.

To be fair, the drug industry, often in partnership with major economic players such as oil companies, is actively involved in research, diagnostics, counselling and treatment in places where the pandemic is most severe. Other large corporations — such as mining company Anglo American and the brewer Heineken — provide high-quality monitoring and treatment programmes for their employees and, sometimes, for their families.

But for reasons of treatment cost and the scale of effort required, these initiatives have taken a long time to get going. Private companies must urgently be encouraged to increase their efforts to take some of the load off governments.

What's more, employers' responses to the pandemic need to reach beyond the provision of medical assistance. Some of the responses needed are more obvious than others. For example, cement company

LaFarge is changing its transport logistics to avoid having several hundred African truck drivers sitting outside its plants for hours or even days at a time awaiting delivery. A more general problem that needs to be discussed and confronted is the web of sexual relationships that exists within some companies.

There are professional skills in the private sector that have not been significantly tapped. In their enlightened self-interest, private companies could be working with governments to apply their expertise in large-scale logistics and financial management to the mammoth task of constraining the epidemic. But even where money is theoretically available, there are bottlenecks. In the Ivory Coast, for example, a Merck scientist called on the company's internal financial expertise to help unclog some issues with resource allocation, only to be told that it would take eight months.

There are many other impediments to private-sector engagement, particularly AIDS politics and suspicion among public officials and at non-governmental organizations. Business expertise would be more readily brought to bear on the pandemic if public-health officials and other parties could agree among themselves what administrative actions need to be taken to fight AIDS.

Nevertheless, such actions are being identified in many hard-hit countries, and corporations and their skills are sometimes being harnessed to implement them. Some of those involved in Product Red, such as the clothing retailer Gap, already lead by example at their own factories in Africa.

Yet there is a nagging concern among AIDS officials that global attention may be drifting away from the pandemic, at a time when the need to confront it aggressively has never been greater. Businesses, not just in Africa but elsewhere, need to be planning for the battle now, so they can act before it is too late. ■

"Private companies must be encouraged to increase their efforts to take some of the load off governments."

Striving for excellence

A German exercise to foster elite universities began inauspiciously but is a step in the right direction.

The German economy, along with the nation's self-esteem, has been staging a recovery of late after a long period of stagnation. Science seems to be enjoying a parallel renaissance. But structural changes are essential if the country's research system is to remain internationally competitive.

Science has long held a prominent place in German society. However, its institutions and practitioners have sometimes been slow to adapt to changing circumstances — notably the rapid emergence of

biotechnology and molecular medicine, the increased internationalization of science, and the desire of many scientists to be involved in the commercialization of their work.

Nonetheless, the German government has continued to steadily increase its investment in science. Germany has benefited immensely from an influx of scientific talent from central and eastern Europe. And efforts have been undertaken to reverse the brain drain of scientists from Germany, with several initiatives aimed at bringing researchers home from the United States.

The main concern is the weakness of Germany's top universities relative to their rivals abroad. Germany's rigid federal structure and its tendency to treat all its universities on an equal basis have contributed to a situation in which no German research university has established itself in the top global tier. According to a survey by the



Shanghai Jiao Tong University, for example, no German university is among the world's top 50.

The government has responded to this challenge with an 'excellence initiative' that will reward universities with particularly convincing concepts in collaborative research and postgraduate training. The first winners were announced on 13 October, and the outcome disproves fears that political considerations might have influenced the selection.

But even if the initiative succeeds on its own terms, it will not in itself provide sufficient resources to propel the winners into the top tier of global universities. Nor will it remove the structural problems, such as the bickering between the states and the federal government over science and education, that prevent German universities from playing to their strengths.

Nonetheless, the initiative shows that Angela Merkel's coalition government is aware of the need for university reform. At the very least, it gives a handful of universities, institutes and research groups the opportunity to showcase their strengths, and provides a tidy sum of money to improve their standing further.

Additionally, the competition has forced every university to think hard about its strengths and weaknesses. For winners and losers

alike, this will help them nurture promising areas of research and develop necessary collaborations across disciplines. It has injected a palpable spirit of optimism, and a healthy extra shot of competition, into the research system.

The selection of the winners on 13 October nearly became a spectacular failure when some politicians, who felt ignored by the scientific jury, allegedly threatened to call the whole thing off. In the end only three universities made the cut — the Ludwig Maximilians University Munich, the Technical University of Munich and the Technical University of Karlsruhe — rather than the expected five. But the fact that some departments in less prominent universities, such as Würzburg or Bremen, gave the winners a run for their money should be an incentive to those who lost out this time to sharpen their research profiles.

Germany's coalition government and those of its 16 states must continue to coordinate their efforts to develop the universities. There are strong indications on the ground that the dominance of the traditional, all-powerful 'Herr Professor' is fading on German campuses, to be replaced by younger laboratory chiefs with a more progressive outlook on how science should be done. That, too, augurs well for the future of science at German universities. ■

A state of flux

A fresh start beckons for the politics of US science.

The United States is approaching a potential turning point for many disgruntled scientists. On 7 November, voters will go to the polls to elect most of their congressional representatives. Many observers foresee at least a partial shift in power, with the Democratic party regaining control of one or both of the houses of Congress. If that happens, the election could set the stage for some significant changes in US science policy.

It is simplistic to regard the Democrats as more 'pro-science' than the Republicans. It is true that academics usually lean to the left, and that scientists have a long history of liberal activism. But Republicans have a strong track record in boosting funding for science, leading the push to double the budget of the National Institutes of Health and, more recently, proposing an increase for the physical sciences in the federal budget for the next fiscal year (see *Nature* 439, 644–645; 2006). Slogans such as the 'Republican war on science', meant to sum up a host of perceived abuses, do not do justice to the complex relationship between science and each of the two major political parties.

Still, the Republican administration of President George W. Bush has come under plenty of fire for its approach to science. Groups such as the Union of Concerned Scientists have charged it with delaying or tampering with scientific decisions. Prominent examples include the Food and Drug Administration's reluctance to make a decision on the Plan B emergency contraceptive pill, against the advice of its own advisory panels, and White House officials editing parts of an Environmental Protection Agency document dealing with climate change.

Now the opportunity arises for scientific groups to start concen-

trating on how to build better relations with the next administration, which will be elected in two years. As a former adviser to President Clinton suggests on page 751, there are some concrete steps that scientists can take if they wish to influence how the next president approaches science.

Researchers should be considering how their own areas of expertise — be it biosecurity, chemical physics or environmental toxicology — can most effectively be incorporated into national policy. They should identify who obtains positions and chairs on relevant congressional committees and open a dialogue with them. They can also develop relationships with other groups with whom they have common interests.

The United States is blessed with an immensely deep reservoir of scientific talent that can, all else being equal, be applied to furthering the national interest. Over the past few years, some communities of scientists have felt that their input is unwelcome in Washington. Specialists in embryonic stem-cell research, for example, have been denied federal funding for work on any new cell lines. Many have resorted to parallel, administratively awkward private or state-funded laboratories; some have even moved abroad. Weapons physicists and specialists in non-proliferation have lost much of their influence. Climate scientists with mainstream views on global warming have been accused of political bias. These are only a few of the areas in which researchers are feeling frustrated by the current administration and its supporters in Congress.

Political leaders will always seek to use science to further their own ends. The political reception to last week's *Lancet* paper estimating that some 650,000 Iraqi civilians have died as a result of the US-led invasion (see page 728) is just one example of the impossibility of a clean divide between science and politics. But US scientists should now be working constructively to develop an improved relationship with the new Congress — whatever its political composition. ■

RESEARCH HIGHLIGHTS

Silica soufflé

Nature Mater. doi:10.1038/nmat1757 (2006)
Mousse, soufflé or powder? All are on the silica-based materials menu developed by Bernard Binks and Ryo Murakami of the University of Hull, UK.

The researchers devised a series of silica nanoparticles that differ in their affinity for water. They chemically modified the particles' surfaces to give properties that vary from completely hydrophilic to very hydrophobic.

When mixed with water, the most hydrophilic particles stabilized a foam. The hydrophobic particles, by contrast, coated water droplets, turning them into 'liquid marbles', which act like a fine powder. The results are pictured.

Nanoparticles with intermediate properties generated powders, foams or sticky, spongy 'soufflés', depending on the amount of water mixed with them. Applications in food technology and cosmetics beckon.



CHEMISTRY

Antibiotic mixed up

Angew. Chem. Int. Edn doi:10.1002/anie.200603892 (2006)

A potent antibiotic, which is effective against 'superbugs' that have developed resistance to other antibiotics, has been synthesized for the first time.

Platensimycin is produced naturally by the bacterium *Streptomyces platensis* and was shown earlier this year to belong to a new class of antibiotic (J. Wang *et al.* *Nature* **441**, 358–361; 2006).

K. C. Nicolaou and his co-workers at The Scripps Research Institute in La Jolla, California, made a racemic mixture of the compound, meaning that their product contained two mirror-image structures, or enantiomers, of the molecule. Future syntheses may improve on this by giving a single enantiomer.

CELL BIOLOGY

Protein makes a good point

J. Cell Biol. doi:10.1083/jcb.200605012 (2006)

Epithelial cells (pictured right), such as those of the skin, arrange themselves in a honeycomb-like pattern to minimize contact with each other. Researchers in Japan have unearthed a clue as to how such cells manage to maintain their angular shapes.

Masatoshi Takeichi at the RIKEN Center for Developmental Biology in Kobe and his colleagues point the finger at the protein Tuba. They suggest that Tuba controls the shaping of junctions between cells by modulating local activation of CDC42, an

enzyme required for organizing the protein skeleton that supports the cell membranes.

When the team used RNA interference to reduce the activity of the gene for Tuba, the epithelial cells looked "curved and slack".

DEVELOPMENTAL BIOLOGY

Gender switch

Nature Genet. doi:10.1038/ng1907 (2006)

Researchers have discovered a gene in humans that, when mutated, causes embryos with a female complement of sex chromosomes to develop as males.

Giovanna Camerino of the University of Pavia in Italy and her colleagues studied a family in which four brothers each had two Xs instead of the normal male complement (XY). Most female-to-male sex reversals result when the X chromosome inherited from the father accidentally acquires a part

of his Y chromosome, carrying a gene called *SRY*. But the brothers had no copies of *SRY* at all. Instead, they had a mutation in a gene called *RSPO1*.

The team suggests that *RSPO1* normally helps to drive the developing gonad to become an ovary. In its absence, genes promoting the formation of a testis take over.

ASTRONOMY

In good company

Astrophys. J. **650**, L41–L44 (2006)

Swarms of small, faint galaxies are being discovered in the Milky Way's neighbourhood using data from the Sloan Digital Sky Survey. The latest detection, confirmed by astronomers at the University of Cambridge, UK, and their co-workers, with follow-up observations from Japan's Subaru telescope, is a dwarf spheroidal galaxy in the constellation Ursa Major. Dubbed Ursa Major II, it is about 100,000 light years away.

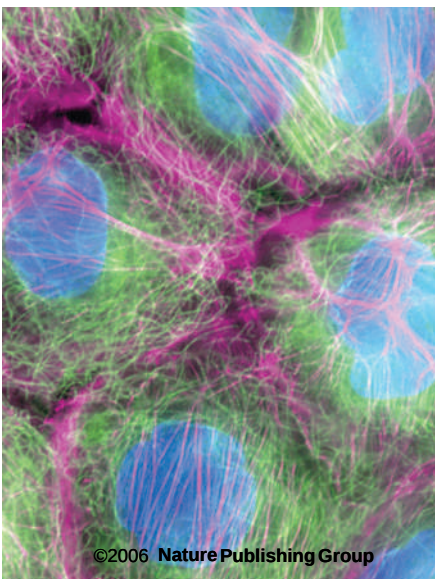
Simulations of how galaxies form in the presence of dark matter predict that many such dwarf galaxies will exist as companions to their bigger cousins. Discoveries like this one are closing the gap between the number of such galaxies known and the much larger number expected if the models are correct.

CRYPTOGRAPHY

They do it with mirrors

Phys. Rev. Lett. **97**, 140502 (2006)

Cryptography's thorniest problem is how sender and receiver can securely share the 'key' — the string of 1s and 0s — they use to encrypt their messages. Some people pursue



J. W. SHULER/SPL

quantum solutions; Jacob Scheuer of Tel-Aviv University, Israel, and Amnon Yariv of the California Institute of Technology in Pasadena propose a classical approach that could work over longer distances.

In their scheme, sender and receiver encode bits in laser light that bounces between them. Each party selects one of two mirrors, which have peak reflectivities at slightly different frequencies, to reflect the beam. Knowing their own choice of mirror, one party can work out the other's choice from the returning light's spectrum. The sender's mirror choice gives the 1s and 0s of the key.

The scheme has yet to be tested, but in principle it can be made arbitrarily difficult for an eavesdropper to determine which bit was sent.

BIOTECHNOLOGY

Joined from birth

J. Am. Chem. Soc. **128**, 13030–13031 (2006)

Fusing two enzymes together can improve the efficiency of biosynthesis of useful compounds, say Oliver Yu of the Donald Danforth Plant Science Center in St Louis, Missouri, and his co-workers.

They selected, from three different plants, genes encoding enzymes that are involved in the synthesis of the antioxidant resveratrol — the compound credited with imbuing red wine with health benefits. The researchers then engineered the genes into yeast cells.

Cells in which two genes were placed back to back, so that the resultant enzymes were linked together, were more efficient at making resveratrol than those in which the enzymes were separate molecules. The same trick worked in human kidney cells, implying that it could be useful in engineering human tissues to make such health-promoting substances.

ANIMAL BEHAVIOUR

On the scent

Curr. Biol. **16**, 1956–1961 (2006)

A pheromone that helps newborn rabbits (pictured right) to find their mother's milk can also teach them to respond to new smells, say researchers in France.

The mammary pheromone in rabbit milk typically prompts pups to root for their mother's nipples. Gérard Coureaud of the European Centre for Taste Science in Dijon and his colleagues tested more than 950 rabbit pups to explore the pheromone's potency.

In one experiment, they wafted a chemical odour over two-day-old pups, while exposing them to the pheromone at the same time. Several hours later, pups started seeking a nipple when they smelt the odour alone. The researchers suggest that the pheromone may help pups to learn the scent of their mother.



G. COUREAUD/CNRS

reversible associations of the receptor and its signalling molecules on the two-dimensional cell surface. The researchers developed a mathematical model that captured this complex process.

CELL BIOLOGY

Assessing affinity

Nature Chem. Biol. doi:10.1038/nchembio823 (2006)

Two difficulties in measuring and modelling receptor activation in live cells are tackled in research by Adrian Whitty and his colleagues at Biogen Idec, a biopharmaceutical company in Cambridge, Massachusetts.

One problem is finding a way to modulate the number of receptors to study the effect of receptor concentration. They dealt with this, in experiments on a receptor complex from a group known as the GDNF family, by adding different quantities of an antibody that blocks free receptors.

The next hurdle is to understand how the receptor is activated, requiring models of the

IMMUNOLOGY

Enzymes tell good from bad

Science doi:10.1126/science.1132998 (2006)

An enzyme that helps cells to detect viral infection responds not to the double-stranded RNA produced when the virus replicates, as previously thought, but to the single-stranded RNA of the virus's genome.

This is the conclusion that Caetano Reis e Sousa of Cancer Research UK, London, and his colleagues reached after studying how the RIG-I enzyme interacts with the influenza A virus. They found that RIG-I recognizes viral genomes that are 'uncapped', meaning they lack the end structures usually found on RNAs made by the host cell. This may have evolved as a way for the immune system to tell self from non-self.

The team also showed that a protein called NS1, produced by the flu virus to suppress its host's immune response, targets RIG-I.

JOURNAL CLUB

Norman H. Sleep
Stanford University, California

A geophysicist sees the positive effects of flood basalt eruptions.

Some people blame massive outpourings of lava for mass extinctions in Earth's history. I have been sceptical of the idea that such events could cause deadly climate change, but quantifying what happened during flood basalt eruptions is the only way to be sure.

Stephen Self of the Open University, Milton Keynes, UK, and his colleagues have done this for one episode in the Deccan province, India. It began to vent around 66 million years ago, just before the Cretaceous/Tertiary (K/T) extinction (S. Self *et al.* *Earth Planet. Sci. Lett.* **248**, 518–532; 2006).

The K/T extinction is linked to an asteroid impact, but some argue that the Deccan eruptions contributed to its beginnings.

Self's team compared the Deccan rocks with the more recent flood basalts in the northwestern

United States and modern basalts in Iceland. They conclude that individual eruption events were huge compared with volcanic eruptions that are known to have affected climate, spewing as much as 10,000 cubic kilometres of lava in a decade. The team also estimated outputs of carbon dioxide and sulphur dioxide.

Carbon dioxide outgassing was insignificant — less than current industrial rates. On the other hand, intense fire fountains carried huge quantities of sulphur dioxide into the stratosphere, blotting out

sunlight globally and triggering volcanic winter.

Volcanic winter would wipe out some species. But a mass extinction? I think not. The episodes recurred over a long enough time period — roughly a million years — for evolution to act.

I wonder whether the Deccan eruptions might even have diminished the effects of the K/T impact, which would have thrown huge amounts of dust into the atmosphere, by 'preadapting' organisms for unexpected cold.

NEWS

Iraqi death toll withstands scrutiny



AU YUSSEF/AFP/GETTY

The uncertainty of figures relating to before the war, and the present security situation, hamper efforts to estimate the Iraqi death rate.

It is one of the most politically charged questions that any researcher can tackle: how many people have died in Iraq since the US-led invasion?

Four public-health experts provided an answer last week. Their result — a death rate that has risen from 5.5 per thousand per year to 13.3 — implies that since the invasion there have been 650,000 'excess' deaths, 2.5% of the

population. Predictably, the finding has met with criticism from supporters of the war from the US president downwards.

If it holds, the study will be a key publication in the growing field of conflict epidemiology. Understanding who is dying, how and where in Iraq is vital for efforts to rebuild the country. So once the political criticisms are set aside, how does the study stack up scientifically?

Estimates of deaths in Iraq all suggest that the death rate has risen since the invasion. But they have been much lower than the latest figure. Iraq Body Count, a left-leaning website that compiles deaths from media reports, quotes a maximum of just under 49,000. US government officials have given figures of 30,000–50,000. And a household survey, conducted in May 2004 by the United Nations and published last year, concluded that the war had caused at least 18,000–29,000 deaths, mostly from violent causes. The only previous estimate of the same order of magnitude is a figure of around 100,000 excess deaths in the first 18 months of the conflict, published in 2004 by Gilbert Burnham and his colleagues at Johns Hopkins University in Baltimore (L. Roberts *et al. Lancet* **364**, 1857–1864; 2004), the same researchers who are behind the new work.

The authors say their estimate is so much higher because their methodology is more comprehensive. They organized detailed interviews, carried out by Iraqi researchers from the Al Mustansiriyah University in Baghdad, in more than 1,800 households in 16 of Iraq's 18 administrative regions. Using random sampling similar to that used in political opinion polls, they documented 82 deaths in the period before the invasion, and 547 during the conflict. More than 90% of these were confirmed by death

Counting the cost of war

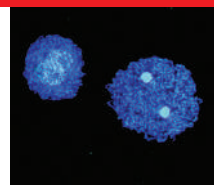
The Iraq study is just one of many that are helping governments and aid agencies understand and respond to the impact of conflicts. Jana Asher of the Science and Human Rights Program of the American Association for the Advancement of Science in Washington DC has produced figures on which armed groups were responsible for deaths in the civil war in Sierra Leone in the 1990s. She says that such studies have become increasingly important over the past 15

years. The growth is due to improved methodology, and to the way that truth and reconciliation committees have come to be seen as an important way for nations to come to terms with a violent history.

Chile set up an early truth and reconciliation committee in 1990, but the concept came to prominence in 1995, when a committee was set up in South Africa. Early committees generally collected qualitative evidence, such as eyewitness statements, says Asher. But the amount of data collected

has increased vastly, and quantitative studies are now being seen as part of a committee's work.

Recent statistical studies include work on Darfur and East Timor. Bodies such as the International Criminal Court in the Hague, the Netherlands, can use the evidence from such surveys when prosecuting war criminals, such as cases relating to Sierra Leone. Most studies are uncontroversial, but the Iraq work is an exception because of the political debate over the US invasion. **J.G.**



SMALLEST GENOME CLOCKS IN AT 182 GENES

How much can you remove from a bacterium before it stops working?

www.nature.com/news

certificates. They show a 95% probability that the death toll has been between 390,000 and 940,000, most of which were due to violence such as gunshots (G. Burnham *et al. Lancet* doi:10.1016/S0140-6736(06)69491-9; 2006).

Data from other conflicts show that such sampling is much more accurate than media reports, which usually account for no more than 20% of deaths. “Random counts force you to go to places that aren’t convenient,” says Jana Asher, a researcher with the Science and Human Rights Program of the American Association for the Advancement of Science in Washington DC. “The media don’t wander off to distant locations. It’s a very different type of data collection.”

Death tolls from Iraqi health officials, the source of US government figures, are also suspect. The Johns Hopkins team says that the process for issuing death certificates still works well in Iraq, but the system for monitoring the number of certificates issued does not. Even before the war, note the researchers, the government’s surveillance system captured only one-third of all deaths.

Yet despite the weakness of other measurements, the new figure has still surprised researchers. Perhaps the most significant con-

cern is the baseline rate for pre-invasion death rates used in the new study. The latest survey, which included questions about the situation before the invasion, put this at 5.5 deaths per 1,000 people per year, in line with figures on Iraq from the US Census Bureau. Iran, which has a well-run health system, has a similar rate, but Iraq was at the time suffering from years of sanctions. Some sources, including the United Nations Population Division, list a pre-invasion figure of 9.7.

“Random counts force you to go to places that aren’t convenient. The media don’t wander off to distant locations.”

The discrepancy does not invalidate the new result, and if the researchers underestimated the pre-war death rate, it’s possible that they may have also underestimated the post-war rate. But some researchers say the paper should have addressed the issue. “There should have been more introspection,” says Beth Osborne Daponte, a demographer at Carnegie Mellon University in Pittsburgh, Pennsylvania. “That increased my discomfort.”

Other researchers share that discomfort. Debarati Guha-Sapir is director of the Centre for Research on the Epidemiology of Disasters in Brussels. She has some methodological concerns about the paper, including the use of local people — who might have opposed the occupa-

tion — as interviewers. She also points out that the result does not fit with any she has recorded in 15 years of studying conflict zones. Even in Darfur, where armed groups have wiped out whole villages, she says that researchers have not recorded the 500 predominately violent deaths per day that the Johns Hopkins team estimates are occurring in Iraq.

But overall Guha-Sapir says the paper contains the best data yet on the mortality rate in Iraq. And none of the experts contacted by *Nature* said that their doubts fatally undermined the study. Some, such as Daponte, would have liked the authors to have better assessed their method’s shortcomings before releasing a result with such political impact. But most say the result is a welcome addition to conflict epidemiology, which is now seen as playing a central role in assessing the severity of wars, and in helping states recover from them (see ‘Counting the cost of war’).

Burnham says he would now like to study patterns of migration in Iraq and the state of the health system. He would also like to estimate deaths based on a sample 4–5 times bigger than that used so far. But survey teams are in danger on the streets of Iraqi towns, and Burnham doubts whether the need for more detailed data justifies the risk. ■

Jim Giles

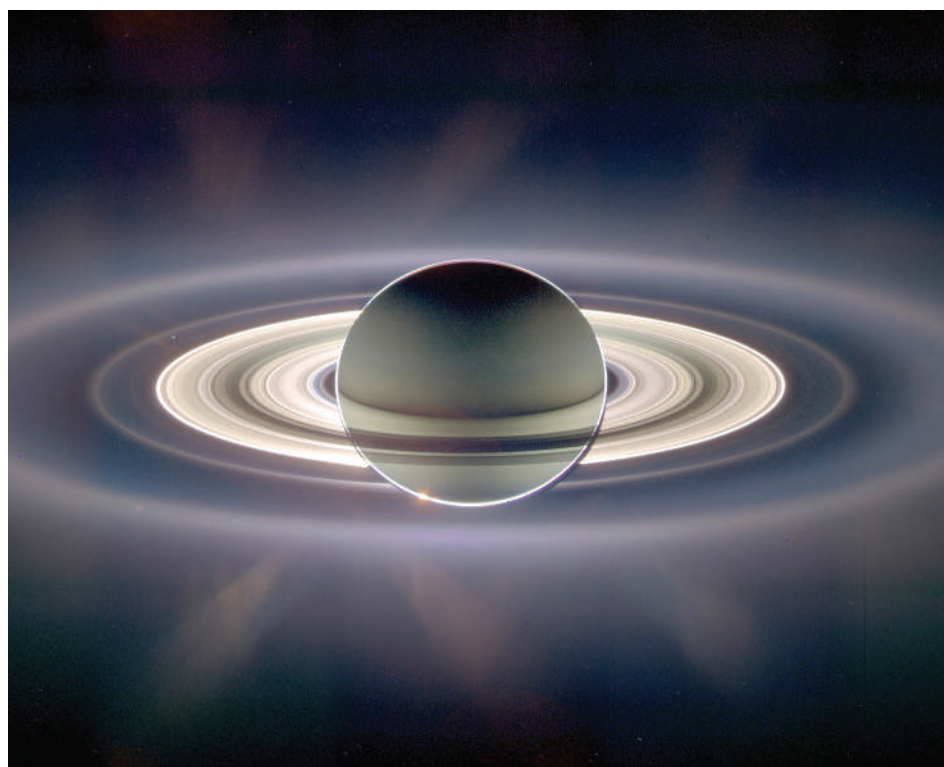
SNAPSHOT Light show

September brought a fantastic photo opportunity for Cassini, the spacecraft in orbit around Saturn. Its trajectory brought it into the ringed planet’s shadow for about 12 hours on 15 September, giving it a striking view with the Sun lying directly behind.

With the glare of the Sun blocked, Cassini’s cameras could pick up light glinting from particles just 1–10 micrometres across. The effect in this exaggerated-colour image is similar, says Cassini imaging scientist Joe Burns of Cornell University in New York state, to the way that stray hairs are highlighted when someone is illuminated from behind.

The light show revealed several ringlets that haven’t been seen before, the team reported at the American Astronomical Society’s Division for Planetary Sciences Meeting in Pasadena, California, this week. One particularly bright ringlet, they say, may be the remains of a recently shattered moon.

Nicola Jones



NASA/JPL/SPACE SCIENCE INST.

Is ice on the Moon just a mirage?

If hopes of colonizing the Moon are ever to become reality, water needs to be found there. Sizeable amounts of ice could provide Moon-dwellers with drinking water and, if they split it into its atomic components, hydrogen for fuel and oxygen to breathe. If lunar water is scarce, colonizers would have to carry water and fuel with them from Earth, or perhaps extract hydrogen from the rocks, either of which could be hugely expensive.

But a paper published in *Nature* this week (see page 835) strengthens the argument that radar data previously interpreted as showing ice could reveal nothing more than rough ground.

The Shackleton crater at the Moon's south pole is a good place to look for deposits of water ice, because parts of it are in permanent shadow from the Sun and so perpetually frozen. Areas of the crater's rim are in permanent sunlight; together these factors make it a good candidate for a lunar base.

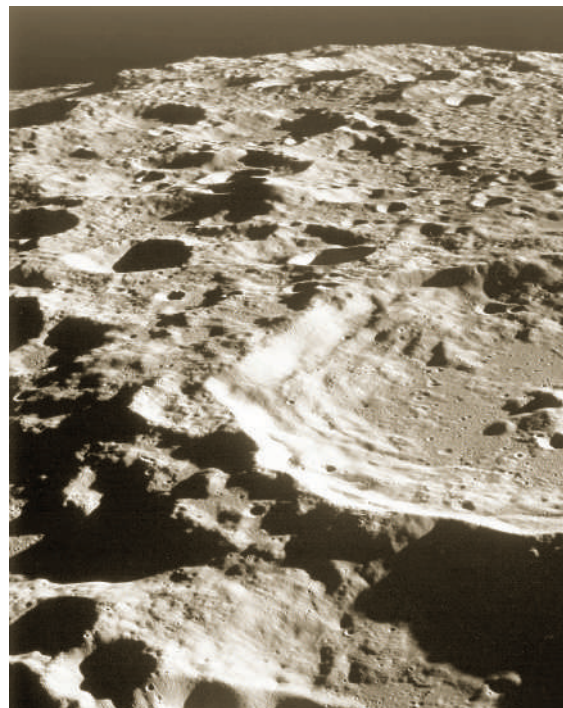
But scientists have argued for a decade about whether there is really ice in the crater. The evidence for Shackleton containing ice deposits is based on the fact that radar signals reflected off the crater are polarized the same

way as the incoming pulse (called a high circular polarization ratio, or CPR) — a property of ice. Rough surfaces have the same effect, but the Moon has few rough areas, because apart from young features such as recent impact craters, the surface has been sandblasted to a fine dust.

But astronomer Donald Campbell, of Cornell University in Ithaca, New York, has never bought this argument. Now he has gathered the highest-resolution radar images made of the Moon so far, and claims that they dash hopes of finding large lumps of ice at Shackleton.

Campbell and his colleagues sent radar signals from the Arecibo radio telescope in Puerto Rico to the Moon, and captured the reflections using the Robert C. Byrd Green Bank Telescope in West Virginia. Campbell says that Shackleton shows the same pattern of high CPR as the similarly shaped Schomburger crater — which is dried out by sunlight. Campbell also found similar patterns in the walls and floors of other young craters. So, he argues, the signal from Shackleton is probably caused by rough rocks.

Campbell did see some evidence of ice at Shackleton, but as dispersed particles in



loose rock, not as lumps. He urges those planning Moon missions to take his findings into account. "Planning for future bases on the Moon should be predicated on there being low concentrations of water ice," he says.

Predictably for such a long-running debate, not everyone agrees. "Don sees one thing and says 'no ice'. I see the same thing and say 'maybe ice,'" says planetary geologist Paul Spudis of the Lunar and Planetary Institute

Missing results might have rung warning bell over trial drug

The company whose experimental drug left six volunteers critically ill failed to provide regulatory authorities with relevant data before the trial, claim two experts. They say a checklist could help to ensure more rigorous risk assessment for future applications, especially given the increasing popularity of drugs that act on the immune system.

The antibody, called TGN1412, activates the immune system's T cells, and was being developed by the German company TeGenero. But in the first human tests on 13 March, it triggered a dangerous over-reaction of the immune system called a cytokine storm. The agency responsible for approving the trial,

the UK Medicines and Healthcare Products Regulatory Agency (MHRA), concluded that this was caused by an unexpected biological effect of the drug.

Now two drug-safety specialists who have investigated the case say relevant information about TGN1412 that might have raised a red flag was missing from the file TeGenero gave the MHRA and other authorities (M. Kenter and A. Cohen *Lancet* 368, 1387–1391; 2006). "If assessors had known, they probably would have started asking questions," says Marcel Kenter, executive director of the Central Committee on Research Involving Human Subjects in The Hague, Netherlands, which also

assesses drugs for clinical trials. His co-author is Adam Cohen of the Centre for Human Drug Research in Leiden.

Kenter and Cohen identified several key pieces of data that were absent from TeGenero's file. For example, they say that tests in mice given a humanized immune system had shown that the antibody caused T-cell numbers to decline, perhaps because they were so fired up they committed suicide. These data were published online in early March by researchers, including those at TeGenero (N. Legrand *et al. Blood* 108, 238–245; 2006).

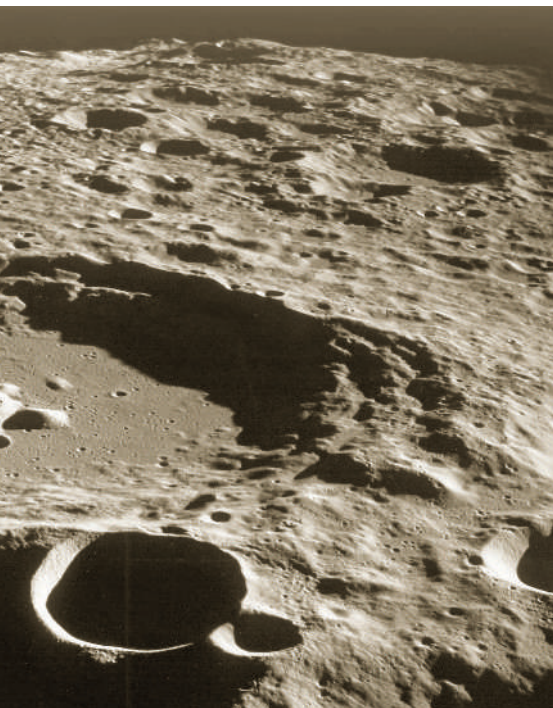
The authors also say the company failed to mention key differences

between the amino-acid sequence of the antibody receptor in rhesus monkeys on which the drug was initially tested and the human sequence. The differences might



Six volunteers almost died during the TGN1412 trial.

B. STANSALL/AFP/GETTY



NASA

On the rocks: astronomers might have mistaken rough craters for lunar ice deposits.

in Houston, Texas. Spudis has believed there is ice on the Moon since the 1994 Clementine mission found high CPR in dark polar regions of the Moon. In 1998, spectrometers on NASA's Lunar Prospector found high levels of hydrogen in the same areas, again suggesting the presence of ice.

have suggested that the antibody would bind more strongly to human than to monkey cells, and perhaps cause a more severe reaction.

Thomas Hanke, chief scientific officer at now-insolvent TeGenero, says all relevant information was given to the MHRA before the trial. Although some of TeGenero's studies were carried out in rhesus monkeys, he says the company used cynomolgus monkeys for toxicological studies because the antibody-receptor sequence of the cynomolgus monkey is identical to the human one.

The MHRA says that at the time of TeGenero's application, senior assessors judged the data submitted to be adequate. But

in a preliminary July report into the incident, an expert scientific group recommended that rules for such phase I clinical studies be tightened to improve the safety of volunteers.

The Paul Ehrlich Institute in Langen, Germany, had also approved an application from TeGenero to carry out a similar trial. Its president, Johannes Löwer, says assessors there did ask TeGenero for additional information during the approval process and received it. "I'm not sure [a company] is obliged to include these data but they did provide them, so they were thinking about the their relevance," he says.

Kenter and Cohen recommend that regulatory

Spudis disagrees that the radar signals from Shackleton and Schomberger craters are the same: "Those patterns look different to me." The CPR signal from Schomberger weakens as the crater deepens, consistent with roughness at its rim, Spudis says, whereas the signal at Shackleton starts halfway down the crater and continues deeper in, as might be expected from ice deposits (see Figure 2, page 837).

Spudis is involved with NASA's Lunar Reconnaissance Orbiter (LRO), set to launch in October 2008, which aims to provide a more detailed map of the Moon. In April, NASA announced that the mission would also include a probe that, by crashing into the Moon, might clear up the ice debate. The probe will split in two, with cameras and a spectrometer on the upper stage recording and analysing the plume of material produced by the lower stage's impact with the Moon. The rest of the probe will crash into the Moon 15 minutes later.

A strong signal of water vapour from the impact could confirm the presence of ice. But an absence of water vapour is unlikely to end the debate, as Spudis argues that the probe could simply hit a dry patch.

The only way to find out for sure will be to land on the Moon and take a sample, says Spudis. Bernard Foing, project scientist on the European Space Agency's Smart-1 lunar orbiter — which ended its three-year mission by crashing into the Moon on 3 September — agrees, and is lobbying hard for the agency to develop a lunar lander that could be launched by 2012. ■

Katharine Sanderson

bodies such as the MHRA adopt a checklist during risk assessments to ensure that they have all the information on a drug's mechanism of action and the effects of human exposure to similar agents. Löwer agrees, and says the Ehrlich Institute already operates a similar checklist, but that other agencies, with less expertise in basic research, might not.

TGN1412 is thought to be an exceptional case. Even so, Kenter says the lessons are particularly important because of the increasing number of drugs that are designed to work specifically in humans and so may have side effects that are difficult to predict from animal trials. ■

Helen Pearson

JOHNS HOPKINS UNIVERSITY MASTERS DEGREE

- Open Houses Oct. & Nov. RSVP Online Now.
- Evening and Weekend Classes
- Earn Your Degree Onsite and Online
- Conveniently located in Washington, D.C.
Baltimore, MD
Rockville, MD
Online.



**MS in Environmental
Sciences and Policy**

MS in Biotechnology

MS in Bioinformatics

**MS in Biotechnology/
MBA Joint Degree**

**MS in Bioscience
Regulatory Affairs**

**Certificate in
Biotechnology Enterprise**

**LEARN MORE AND
APPLY ONLINE**

**advanced.jhu.edu
1.800.847.3330**





ONE GENE BETWEEN DOGS BIG AND SMALL?

Size study highlights possibilities of the dog genome.

www.nature.com/news

ALAMY

Kudos, not cash, is the real X-factor

Since the success of the initial Ansari X Prize competition to create a reusable private spacecraft, technology competitions are popping up everywhere, offering millions of dollars in prizes. So can such financial incentives really revolutionize a field? Supporters say they can, but forget about the money — the real prize is publicity.

This week's X Prize Cup, scheduled to begin on 20 October, is the latest instalment. The two-day festival in the New Mexico desert will feature three competitions, including a NASA-sponsored lunar-lander contest and more than US\$2.5 million in prize money.

NASA has issued a series of challenges to improve everything from elements of the proposed space elevator to the gloves that astronauts wear in space. And the US House of Representatives recently passed a bill to create the H Prize, complete with a \$10-million reward, to encourage the development of technology for hydrogen-powered vehicles. "We were directly inspired by the X prize," says Congressman Bob Inglis (Republican, South Carolina), who sponsored the H Prize bill.

Meanwhile, the X Prize Foundation has announced its own latest challenge, offering \$10 million to the first private team to sequence 100 human genomes in 10 days. The Archon X Prize for Genomics, launched on 4 October, is meant to stimulate technology that will bring down the cost and time of sequencing, accelerating the development of personalized medicine tailored to individual genomic attributes.

As with the first X prize, those chasing the genomics award will probably invest much more than \$10 million. The true pot of gold is not cash, but publicity and a head start at grabbing a slice of a growing market.

"The prize is really not the X prize," says Steve Lombardi, senior vice-president of marketing at Helicos BioSciences, a sequencing company based in Cambridge, Massachusetts. "It's that this is going to put these sorts of ideas in front of the general public." Inglis agrees, adding that he hopes private contributions will boost his H prize to \$50 million: "I think having a more substantial prize makes it more interesting to the public."

The details of the latest X prize diverge somewhat from the model established by the first competition. That was created to fill a clear hole in the private sector, but personalized medicine has been receiving media attention for years. The genome-sequencing industry is littered with companies large and small, and venture capital has already shown an interest



X marks the spot: the first private team to sequence 100 human genomes in 10 days will win \$10 million.

in advanced sequencing technology. For the past two years, the US National Institutes of Health (NIH) has been funding 'revolutionary' sequencing programmes with the aim of one day developing technology to sequence individual genomes for only \$1,000.

Geniuses and mavericks

Kathleen Wiltsey, executive director of the Archon X prize, acknowledges that a sequencing industry already exists, but says it still needs the stimulation the prize would bring: "Given the scale of the task and all the directions this can take, the industry is really underscaled for the demand."

The prize will also capture the interest of people who weren't necessarily thinking about sequencing before, she adds. "The idea of the prize draws a lot of attention. It draws geniuses and mavericks out of the woodwork."

George Church, director of the Lipper Center for Computational Genetics at Harvard Medical School in Boston, Massachusetts, who is on the advisory board for the Archon prize, thinks the winner will probably be someone from an established company that has deep pockets. "Everybody who's serious about it is already getting NIH money or venture capital," he says.

But some are more optimistic about attracting attention. "I think it will encourage people to enter this field," says Amit Meller, a bioengineer at Boston University who has been trying for nine years to use nanotechnology to speed up sequencing. "There are many others out there who are working quietly, and now they may start to be more aggressive."

One point everyone agrees on is the power of the X-prize model to focus public attention on the social and political issues surrounding a field. For the H prize, that means energy policy, says Inglis; for the latest X prize, it's how to interpret and protect genomic information.

Francis Collins, director of the US National Human Genome Research Institute, has seized this opportunity to voice his concerns about protecting the privacy of genomic information. Genomics entrepreneur Craig Venter, on the other hand, has used his time in the spotlight to talk about the need to educate the public about the fallacy of genetic determinism.

Church adds that other industries will also start thinking about how to capitalize on the coming flood of genomic sequences. "This will make downstream users start thinking about it ahead of time," says Church. "It's like having a contest for Internet browsers back in 1992." ■

Heidi Ledford

A. WONG/GETTY/NEWS.COM



T. EISNER

DO IT YOURSELF

Ecologist Thomas Eisner is a photographer as well as a researcher. But the effects of Parkinson's disease have led him to explore a new way to capture images, using a colour photocopier. Eisner lays objects such as flowers and shells (as shown here) on the photocopier stage, covered with a black cloth — ideal for anyone with limited mobility, he suggests.

ON THE RECORD

“Science should claw back its nineteenth-century glamour.”

Columnist Simon Jenkins argues on *The Guardian's* Comment is Free website that specialist science education is not for the masses.

“How depressing to see ignorance championed in this way.”

Response to Jenkins's article from an anonymous poster.

SCORECARD

Mushrooms
After spinach and then lettuce were withdrawn in the United States owing to fear of bacterial contamination, one online betting site has made mushrooms the favourite to be the next foodstuff to be recalled, at 6/1.

Cloned cats
Genetic Savings & Clone, a biotechnology company that sold cloned pets, is to close due to lack of demand. Since the company opened in 2000 it has created five cloned cats but sold just two.

NUMBER CRUNCH

2–3 hurricanes were estimated to make landfall in the United States this hurricane season.

0 hurricanes have made landfall so far.

\$6.5 billion has been lost by hedge fund Amaranth, which had bet on hurricanes hitting US oil depots and forcing up fuel prices.

Sources: New York Times, NOAA

Japan's new premier chases innovation

TOKYO

In a departure from tradition, Japan's new prime minister, Shinzo Abe, has appointed a vocal supporter of science, Kiyoshi Kurokawa, as his special adviser. The two men promise to reform dramatically Japan's rigid structures of scientific education, funding and decision-making, in order to boost technology and innovation in the country. But with hard details lacking, some observers are left wondering how much will really change.

Abe took over the premier job from Junichiro Koizumi in late September. He is expected to continue Koizumi's reforms of science policy, including making universities more competitive and increasing career opportunities for young researchers and women.

But in a surprise move, on 3 October Abe appointed Kurokawa, former president of the Science Council of Japan, as the cabinet's special adviser. Kurokawa is the first such adviser with a science background — his four predecessors were economic or legal experts. The motive seems to be a desire to boost the economy without falling back on tax cuts, which are unpopular in Japan. “I think the prime minister is aware that he needs a scientific adviser,” says

Kurokawa. “To keep growing, we need innovation in science and technology that can totally change society, like the Industrial Revolution and the Internet.”

Hiroyuki Abe, a member of the Japan Council for Science and Technology Policy and former president of Tohoku University, agrees, adding: “Kurokawa is the best choice because he's been calling for innovation and reforms at the top of academia.”

“To keep growing we need innovation that can totally change society, like the Industrial Revolution and the Internet.”

Among other things, Kurokawa will help to create Innovation 25, a plan to forecast what society's needs will be in 2025, and suggest what research should be done in medicine, information technology and the environment to meet those needs. The push

gives greater impetus to the direction Koizumi wanted to move in while he was in office. A team led by Kurokawa is expected to complete the plan by around June next year. It won't include any detailed policy recommendations, but could pave the way for increasing funding and overhauling the grant system, while reforming education and research systems to make them more flexible.

Ikuo Kabashima, professor of political science at the University of Tokyo, says that

SHINZO ABE, 52, PRIME MINISTER

A prince in the political world. His father was foreign minister, and his grandfather was Prime Minister Shinsuke Kishi. The first prime minister born after the Second World War, Abe held key positions in Koizumi's administration. But he is more hawkish than his predecessor, and aims to change, or at least reinterpret, Japan's pacifist constitution to give the self-defence forces a more assertive role. He also promises to improve political relations with China and South Korea. Keen to revive the public education system, his catchphrases are “rechallenge” (giving people second chances to win good jobs) and, of course, “innovation”.



K. SASAHARA/AP

KIYOSHI KUROKAWA, 70, SPECIAL ADVISER

A physician specializing in kidney medicine. Trained in the United States, he practised and taught there for 15 years before serving as a professor at both the University of Tokyo and Tokai University. Until September he held the position of president of the Science Council of Japan, where he reformed its structure partly by increasing the number of female members and by changing the way members are selected. Kurokawa has been outspoken in criticizing Japan's universities and researchers for boxing themselves into narrow research fields, and not considering international collaborations or how their research could be of wider use.



TWP/PHOTO/CORBIS

**ELEMENT 118 MADE
IN THE LAB**Three atoms made for less
than a millisecond each.www.nature.com/news

changes could start coming into effect once Shinzo Abe has ensured the stability of his administration after an upper-house election scheduled for the summer next year. "Calling for innovation as a drive for economic growth sounds better to the public than tax cuts," he says. But he warns that conservative ministries and universities are likely to put up quite a fight rather than lose their current powers.

Researchers will also take some convincing about the reforms. Fujio Masuoka, a professor at Tohoku University and the inventor of flash memory, is sceptical about the likelihood of reform. He has experienced rejections from various ministries over the past few years, when applying for funding to create a three-dimensional semiconductor chip. He gave up three years ago and started working with a company in Dubai instead, which invested ¥10 billion (US\$84 million) in his project. "Whenever we want to try new things, we hit a wall because a consensus of several reviewers is needed and that system makes it hard to evaluate unique research properly," says Masuoka. "I don't expect anything to change in Japan." ■

Ichiko Fuyuno

Currency of science: technology innovation is seen as central to Japan's economic growth.



eppendorf advantage ✓

www.eppendorf.com/advantage
Get Eppendorf quality today!

eppendorf® is a registered trademark.

Promotion
September 1 –
December 31,
2006

Take advantage now!

With **Eppendorf Advantage** offers it is now easier than ever to add Eppendorf premium quality products to your lab.

Spin & Mix!

Introduce the new spirit of centrifugation into your laboratory with the new **Centrifuge 5424!** Or say “hello” to **MixMate**, our new all-around mixer for small volumes, and take advantage now for a chance to win a digital camera! Great offers* are now available from participating Eppendorf Dealers.

Our exciting promotions are just a few clicks away!
For more details visit www.eppendorf.com/advantage

*Offers may vary by country.

FDA urged to assess cloned animals like drugs

Several advocacy groups last week petitioned the US Food and Drug Administration (FDA) to regulate cloned animals as “new animal drugs” under the Federal Food, Drug, and Cosmetic Act, the benchmark law governing the approval of ordinary drugs.

The groups argue that the law’s broad definition of ‘drug’ supports their claim that a clone is a drug, and thus that investigations into its safety should be the same as those used for drugs.

“The FDA has had this issue before them for several years now, and hasn’t acted,” says Jaydee Hanson, a spokesman for the Center for Food Safety in Washington DC. The petitioners are also asking the FDA to make compulsory its 2003 request to companies not to sell foods from cloned animals.

In 2003, the FDA issued a draft risk assessment which concluded that foods from normal, healthy clones “do not appear to pose increased food consumption risks”. But the agency also said that additional data on the issue would increase its confidence in this conclusion.

Big-headed mouse found on Cyprus is a new species

Genetic tests have confirmed that an unusual type of mouse found on Cyprus is indeed a new species, it was announced last week. The mice were first noted in 2004 for having a normal body size but bigger heads and teeth than mouse species in mainland Europe.

The new species, *Mus cypriacus*, was found by Thomas Cucchi of Durham University, UK, and colleagues. It is believed to have diverged from its mainland counterpart, *Mus macedonicus*, in the Middle Pleistocene era, between 650,000 and 350,000 years ago (T. Cucchi *et al.* *Zootaxa* 1241, 1–36; 2006).

Keith Dobney, a bioarchaeologist and colleague of Cucchi’s at Durham University,



Keep your eyes open: the discovery of a mouse raises hopes of finding more mammal species.

Tortoises not so lonesome on Galapagos island

The remote island of Pinta in the Galapagos archipelago is to be the setting for a conservation move unprecedented in the history of these iconic islands.

Last week, researchers announced that 100 giant tortoises from Española — another island in the Galapagos — will be tagged with microchips and set free on Pinta's volcanic slopes later this year. They will replace Pinta's own subspecies of tortoise, only one of which survives: the celebrated Lonesome George (pictured), who



was discovered and taken into captivity on the island of Santa Cruz nearly 35 years ago.

It will be the first time conservationists in the Galapagos have used 'taxon substitution' — using one taxon as a stand-in for another. The move should restore some balance to Pinta's ecology and strengthen efforts to conserve the Española tortoise, a subspecies that at one time was down to just 15 individuals but is now back in the thousands.

notes that there may be still more new species to discover in Europe. "Perhaps not large animals," he says, "but there certainly is a chance of finding these small mammals. This one was right under everyone's noses."

Cell biologist takes on PNAS editorship

The *Proceedings of the National Academy of Sciences* has named a new editor-in-chief, Randy Schekman, a cell and developmental biologist at the University of California, Berkeley.

Schekman replaces Nicholas Cozzarelli, also of Berkeley, who died of lymphoma in March after editing the journal for 12 years.

Originally trained as a biochemist, Schekman says he hopes during his four-year term to broaden the research topics published (particularly from the physical sciences), "tighten up" the review and decision-making process for articles, and compete more with other major journals for manuscripts.

"The biggest challenge is to make sure the journal attracts the most important manuscripts," he says. "It is a big job."

Pressure group pushes for 'nano-hazard' symbol

A campaigning group that has previously called for a moratorium on nanotechnology research is now pushing for the creation of a 'nano-hazard' symbol. The ETC Group, based in Ottawa, Canada, last week launched a competition to select a design.

One suggestion already on display at www.etcgroup.org/nanohazard is a carbon 'buckyball' inside a yellow warning triangle — resembling the 'nuclear hazard' sign.

Ann Dowling, who chaired the working

group on nanotechnology of Britain's Royal Society and Royal Academy of Engineering, notes that much nanotech work poses no new safety risks, so "a universal 'nano-hazard' symbol would be wholly inappropriate". But she acknowledges concerns about the safety of 'free' nanoparticles: "We do need a discussion about labelling products containing these substances so that consumers can choose whether to use them or not."

A longer life when matter meets antimatter

Annihilation has been delayed by allowing a chemical reaction between matter and antimatter to take place in a vacuum.

Protonium, which consists of a proton and an antiproton, is the product of a reaction between a hydrogen molecular ion (which has two protons and one electron) and an antiproton. Researchers normally make protonium by smashing these two molecules together at high speed, and the resulting atom usually annihilates itself almost immediately.

But a different method honed by the ATHENA atom-smashing collaboration based at CERN in Geneva has produced protonium that lives for a millionth, rather than just a trillionth, of a second (N. Zurlo *et al. Phys. Rev. Lett.* 97, 153401; 2006). This longer life was made possible by allowing the hydrogen ion and the antiproton to meet in a vacuum. As a result, they were attracted by their electronic charges, as occurs in a normal chemical reaction.

Correction

Our News Feature on decadal surveys (*Nature* 443, 386–389; 2006) cut a year off the GLAST satellite's development time; it will be launched in late 2007, not late 2006.

eppendorf
advantage ✓



New! MixMate!

Win a digital camera. Details at
www.eppendorf.com/MixMate



*Offers may vary by country.

Promotion
September 1 –
December 31,
2006

Current Eppendorf Advantage promotions include great value packages* with the new Centrifuge 5424:

- A new high-class 24-place microcentrifuge with many new and innovative features!
- Choice of four rotors and two different control panel versions.

For more information contact your Eppendorf Partner today or visit
www.eppendorf.com/advantage

eppendorf
In touch with life

BUSINESS

A firm response to AIDS

'Product Red' is the private sector's bid to fight HIV. But is it too little too late? **Colin Macilwain** investigates.

In the 9 October issue of *New Yorker*, the actress Penelope Cruz is seen reclining across an advertisement sporting a cute red tee-shirt that proclaims the clothing chain Gap's commitment to fighting AIDS in Africa. She and fellow celebrities adorn almost every other page of the magazine, as Gap seeks to make a splash for its 'Product Red' partnership with the Geneva-based Global Fund to Fight AIDS, Tuberculosis and Malaria.

Launched in the United Kingdom in March 2006, the Product Red campaign has raised \$10 million so far. Participating firms agree to contribute a share of profits on a particular product line to the Global Fund. Fund director Richard Feachem says Product Red "has the potential to secure hundreds of millions of dollars per year" as the campaign rolls out across the United States and the rest of the world.

But is the private sector pulling its weight in the fight against HIV? At a 5 October meeting at London's Royal Institute of International Affairs, better known as Chatham House, many participants argued that it is not. "The global business community is doing only a fraction of what it could be on HIV and AIDS," says Michel Sidibe, director of regional support at the Joint United Nations Programme on HIV/AIDS (UNAIDS). "There is an urgent need to turn the widespread rhetoric into action," warns Feachem.

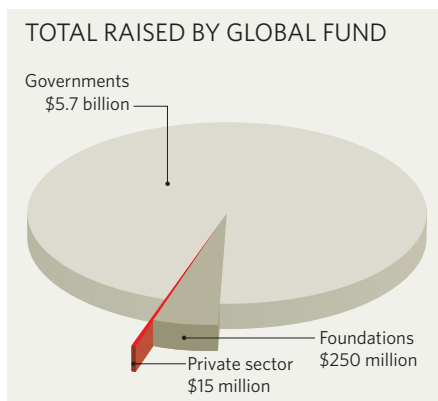
As of last March, Feachem notes, contributions from individuals and companies amounted to a princely 0.1% of the \$5 billion given to the Global Fund since it was established in 2002. The Bill & Melinda Gates Foundation has contributed a further \$250 million. John Tedstrom, director of the Global Business Coalition on HIV/AIDS in New York, says he puts the private sector at the head of the list of those who have to do more. Tedstrom describes the level of business participation at July's international AIDS meeting in Toronto as "terrible".

As well as money, private-sector involvement can take the form of product development and pricing, most obviously by the drug industry; diagnosis and treatment for company workers; the provision of logistical support to those fighting the pandemic; and the marketing of charitable opportunities and awareness-raising, as in Product Red.

Although many officials remain disappointed at the size of the effort, some industry managers are more hopeful. "Businesses around the world are applying their core competencies to the AIDS epidemic," says Jeff



Spreading the message: singer Bono launching Motorola's phone for Product Red in London.



Sturchio, vice-president for external affairs at New Jersey-based pharmaceutical giant Merck. "We should be multiplying these responses — and if we did, that would add up to a much more effective response than we have now."

Partnerships between drug companies, governments and non-governmental organizations are delivering HIV testing and treatment to thousands of Africans in AIDS hotspots. Merck, for example, is active in Côte d'Ivoire, Botswana and Bonny Island, the bustling centre of Nigeria's booming liquefied natural gas industry.

Over the past five years, major multinationals in Africa have started extensive testing and treatment programmes for their staff. Mining group Anglo-American expects to test half of its 120,000 employees in southern Africa this year, says its vice-president for health, Brian Brink. The company estimates that 23% of its staff —

28,000 people — is HIV positive. Antiretroviral drug treatment is available to the 3,800 employees who are in the company's private health insurance scheme, but not to their families.

Some mines already have 90% of their staff taking part in voluntary counselling and treatment programmes, says Brink. "Businesses are beginning to demonstrate that a response to AIDS makes good business sense," he says. If more companies respond, he adds, "it'll take a huge load away from the governments".

The brewer Heineken has extended similar programmes to the families of its workers in Nigeria and elsewhere. Katinka van Cranenburgh, international manager of the brewer's health department, says that 7,000 staff, 5,000 spouses and 20,000 children have access to its healthcare programmes in Africa. Starting the programme was "a dire necessity", as AIDS swept through the company's workforce, Cranenburgh says, and since it started, there's been a "dramatic decrease" in deaths and in time lost to illness.

Private battles

For most African employers and employees, however, this kind of coverage is out of reach. Elvis Musiba, president of the Tanzanian Chamber of Commerce, Industry and Agriculture, runs a company that organizes tours of the Serengeti National Park. In 2003, as the pandemic decimated his drivers, he contacted a consultant and arranged testing for the company's 150 staff, starting with himself.

Musiba complains bitterly that the company had to do this itself, despite reports that Tanzania is getting \$600 million in aid to deal with AIDS. "We asked for assistance and got nothing," he says. "There's a myth that the private sector has the money to handle this on its own."

That is the crux of battles over public-private-sector partnerships on HIV in Africa and elsewhere. The direction of cash flow is often in dispute. As Robin Gorna at the UK Department for International Development tells business leaders: "When you say it makes good business sense, I think: you don't need our money."

For Product Red, public officials note, participating companies demanded 'targeting' of their funds — in this case on AIDS in Africa, rather than the broader remit of the Global Fund — so that they get the kudos. "The only way to get private money into the Global Fund," notes Gorna, "is by earmarking."

Hopes are high for Product Red, though. Since it got going in Britain, with American Express cards, Motorola phones, Armani shades and Gap tee-shirts, "all the feedback has been very positive", says Rajesh Anandan, head of private-sector partnerships at the fund. Anandan explains that the profit-sharing deals on each 'red' product are

"When you say it makes good business sense, I think: you don't need our money."
— Robin Gorna

separately negotiated, to suit items as different as coats and credit cards.

He stresses that the campaign must be accessible. "We're trying to reach hundreds of millions

of consumers in a cause they haven't thought about," Anandan says, adding that subsequent communication on websites and elsewhere will "get to the heart of the issue" and help revive public awareness of AIDS.

What AIDS officials want most of all from the private sector, according to several speakers at the Chatham House meeting, is their expertise in logistics and financial management. Financial managers at Merck helped with the company's partnership programmes, for example, and management consultants Accenture volunteered to help UNAIDS deal with financial bottlenecks that were slowing national AIDS programmes.

But much more help of this kind is needed. Reasons for the shortfall may be mutual suspicion and a clash of cultures between public and private sectors. "In the private sector, they like to act," observes Elizabeth Mataka, director of the Zambia National AIDS Network, "while we have to talk about structures. We spend for ever doing that, when there is actually a lot to do." ■

See Editorial, page 723.

Product Red ► www.joinred.com

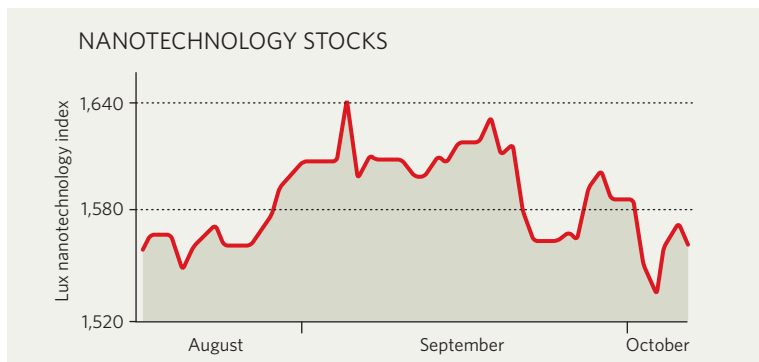
IN BRIEF

VACCINE VALUE US giant Pfizer is buying PowderMed, a vaccine company based in Oxford, UK, for an undisclosed sum. The move reflects burgeoning interest in vaccines among major drug companies in light of the threat of a global flu pandemic, analysts say. PowderMed — which started up two years ago as a spin-off from Chiron, the biotechnology company — is developing vaccines for flu, cancer and AIDS. The vaccines work by using fragments of DNA, blown into the skin. Once the DNA is expressed by cells in the skin, the resulting antigen stimulates an immune response.

AIRBUS AXE Louis Gallois, the new chief executive of Airbus, has warned of "painful" job cuts ahead as the airliner manufacturer adjusts to production problems with its new superjumbo, the A380 (see *Nature* 443, 385; 2006). Reports suggest that up to 10,000 jobs could go at production plants in France, Germany, Spain and the United Kingdom. Airbus has repeatedly delayed delivery dates for its new airliner, with problems said to be connected to the intricate way in which production is divided between the four nations.

BIOTECH BOON Genentech has reported surging profits for the quarter ending in September, with a 58% increase over the corresponding period in 2005. The biotech company's income was driven by US sales of more than \$1.8 billion — up by 34% on the previous year. Genentech, based in South San Francisco, sells high-value anticancer agents based on antibodies, including Rituxan for non-Hodgkin's lymphoma and Avastin for colon cancer. Lucentis, the company's new drug for age-related vision loss, has already logged sales of \$153 million since its launch in June.

MARKET WATCH



It had been a quietly optimistic summer for nanotechnology stocks. But progress was checked in September, with a couple of well-fancied companies suffering reversals.

The Lux Nanotech Index, which measures a basket of smaller companies specializing in nanotechnology and larger ones with interests in its application, ended the third quarter up slightly from August, but well short of its April peak of 1,840.

One company hit was Toronto-based Nucrust, which makes silver nanoparticle wound dressings that have impressed the market. Nevertheless, its share value almost halved after an announcement on 20 September admitting the failure of efficacy trials on NPI 32101, a cream for treating eczema.

Another fashionable nanotech stock on the slide is Minneapolis-based NVE Corporation, a leader in applying

spintronics to sensors and memory chips. After almost touching \$40 on 17 September, the company's stock has slipped back to less than \$28.

British company Cambridge Display Technology fared better, its stock rising sharply after an upgrading by analysts optimistic about the long-term prospects for the company's polymer light-emitting-diode display screens.

According to Peter Hebert, chief executive of Lux Research, the New York-based consultancy that compiles the index, the overall health of the sector is vouched for by a stream of collaborations between small nanotech suppliers and major companies. In August, for example, Turkey's largest oil company, Petrol Ofisi, agreed to work with Oxonica, based in Oxford, UK, to adapt the British company's nanocatalysts as fuel additives.

Colin Macilwain

SHOWDOWN FOR CAPITOL HILL

Can science influence politics in the forthcoming US elections? *Nature* investigates how Democrats and Republicans are striving to win the hearts of voters.

Heather Wilson has something she wants the voters of New Mexico's first congressional district to know about her: unlike President George W. Bush, she supports embryonic stem-cell research. In a local television advertisement last month, Wilson told viewers in Albuquerque and its environs: "The president vetoed the stem-cell bill, and I voted to override his veto because it was the right thing to do."

It is not that surprising for a candidate to say an unpopular president is wrong, or that a popular biomedical cause is right. It is rather more surprising when the candidate and the president are members of the same party. But Wilson, the only woman in Congress who has served in the military, is a moderate Republican in a close fight to keep her seat. The stem-cell issue is one that she thinks may help her in the struggle to a fifth term in the House of Representatives, despite the fact that in other parts of the country some of her fellow Republicans are making opposition to the research a strong part of their campaign.

Most of the time, American voters couldn't care less about science. "If you look at what people are hot and bothered about, it's health care, Iraq, taxes, education, things of that sort," says Daniel Greenberg, a science-policy expert based in Washington DC. "Science and technology — they don't know anything about it." (See 'Q&A', page 744.) As a result, there's not much room for science in the typical US campaign — including the upcoming 7 November mid-term elections, which will decide who holds all 435 seats in the House of Representatives; one-third of the Senate's 100 seats; and 36 governorships. "Science plays a very little role because facts play a very little role," says Tony Massaro of the Washington-based League of Conservation Voters. "Issues, such as they are, are indicators for the values of the candidates."

Some liberal groups think that the Bush administration's record on science can be seen as a reflection of its values. They hope to exploit this, painting a picture of an administration opposed to objective truth and intellectual progress. Over the past six years, activists have assembled a litany of issues where, they say, the Bush administration has either ignored scientific evidence or sought to manipulate it — from delayed decisions on the Plan B emergency contraceptive to altered documents on global warming. This 'Republican war on science', as it was called in the title of a 2005 book by journalist Chris Mooney, has proved a powerful rallying point for scientists disillusioned



with the current administration. And there are many.

But despite the fervour of some of its devotees, there is little evidence that this radical thesis is having any more effect in this race than similar ideas had in the 2004 presidential election. Back then, a group of respected scientists, including two dozen Nobel laureates, publicly accused Bush of "misrepresenting and suppressing scientific knowledge" (see *Nature* 427, 663; 2004) and went on to support John Kerry's presidential bid. This year's race has seen the formation of a political advocacy group called Scientists and Engineers for America, which includes many of the same researchers.

The group says it will work to raise public awareness of perceived abuses of science, and hopes to persuade voters to elect like-minded candidates. But the way the organization is set up constrains it from endorsing any specific candidate in any specific election. And despite charges that it could be

Representative Heather Wilson has distanced herself from the president.



J. SCOTT APPLEWHITE/AP

RUNNING ON SCIENCE

A look at nationwide races reveals three candidates who are challenging incumbents with science-based arguments.



Patricia Madrid, New Mexico

On the surface, the congressional race in New Mexico's District 1 looks like any tight political battle. The two candidates have railed against corruption and taxes, and quibbled over who had the better location in the state-fair parade. But as the challenger tries to take the scientific high ground, science has emerged as a new political weapon.

The race pits Heather Wilson, the Republican incumbent (see main story), against Patricia Madrid, a Democrat and the state's attorney general. In early August, Madrid released a television advertisement highlighting Wilson's campaign contributions from the oil industry, and then announced her own pledge to "invest in alternative energy and fight global warming".

Madrid's climate-change pledge was a sign of the times, says campaign spokeswoman Heather Brewer. "Global warming really resonates with people," she says. "It polls really well." The issue is not new to Madrid — as attorney general, she led the charge for New Mexico to join a coalition of states that sued the Environmental Protection Agency in an attempt to force it to regulate carbon dioxide as a pollutant.

If elected, Madrid says she plans to encourage the development of alternative energy sources by creating tax rebates for individuals and companies that use renewable energy sources.

The only scientific issue discussed during the candidates' 17 September debate was stem cells. Both Wilson and Madrid expressed support for federally funded embryonic stem-cell research, but Madrid took the opportunity to censure President George W. Bush for blocking federal funding for such research. "This administration is anti-science," she declared.

Although scientific issues are part of Madrid's campaign strategy, Brewer cautions that discussions about energy and climate change rarely extend beyond superficial political arguments. "Honestly, I just wouldn't call it a scientific discussion," she says. **H.L.**



Claire McCaskill, Missouri

Top billing on the campaign website of Claire McCaskill recently went to "real Missourians" complaining about incumbent Republican senator Jim Talent, and his failure to support stem-cell research. Here you can watch elderly Ida explain that she has Parkinson's disease. If she could tell Talent one thing, she proclaims in a shaky voice, it would be this: "I hope you and your family will never need the cures offered through stem-cell research."

Such is the political challenge that McCaskill, the Democratic state auditor, has laid for Talent as she seeks to unseat the conservative freshman senator. Talent, a former law professor and long-time politician, voted this summer against a bill that would have lifted constraints on federal funding for human embryonic stem-cell research. Talent called it unacceptable because it "would use tax dollars to fund research that would destroy human life at the earliest stages".

Talent is also opposing a state ballot measure, which McCaskill supports, that has put the stem-cell issue front and centre in Missouri's elections. The initiative would amend the state's constitution to allow scientists to conduct any form of stem-cell research allowed under federal law. It is a response to attempts in the state legislature to criminalize the research.

"There has been a type of warfare declared against science in Missouri," says McCaskill. "This is just about not criminalizing scientists and patients and researchers." The ballot issue has split the Missouri Republican party: former three-term Republican senator John Danforth, whose brother died of Lou Gehrig's disease, is co-chairing the coalition supporting the amendment.

Talent's stem-cell positions play well to his base in this conservative farm state, which, with one exception, has not elected a Democratic senator since 1980. But polls show that the ballot initiative has support among close to 60% of Missourians — a wave McCaskill hopes will tip the balance her way in this very close race. **M.W.**



Jerry McNerney, California

California's 11th congressional district, which sprawls across several counties east of San Francisco, has belonged since 1993 to Republican Richard Pombo — chairman of the House Committee on Resources and perhaps the least-favourite congressman of the League of Conservation Voters. It's also the site of a high-profile showdown between Pombo and his Democratic opponent, Jerry McNerney, a wind-energy engineer with a PhD in mathematics.

McNerney lost to the well-funded Pombo in 2004, and polls again paint him as an underdog in this competitive race. But he carries the fondest hopes of many US environmentalists, who are worried by Pombo's proposed radical reshaping — gutting, they say — of the Endangered Species Act and his plan to introduce oil drilling in areas that are currently off-limits.

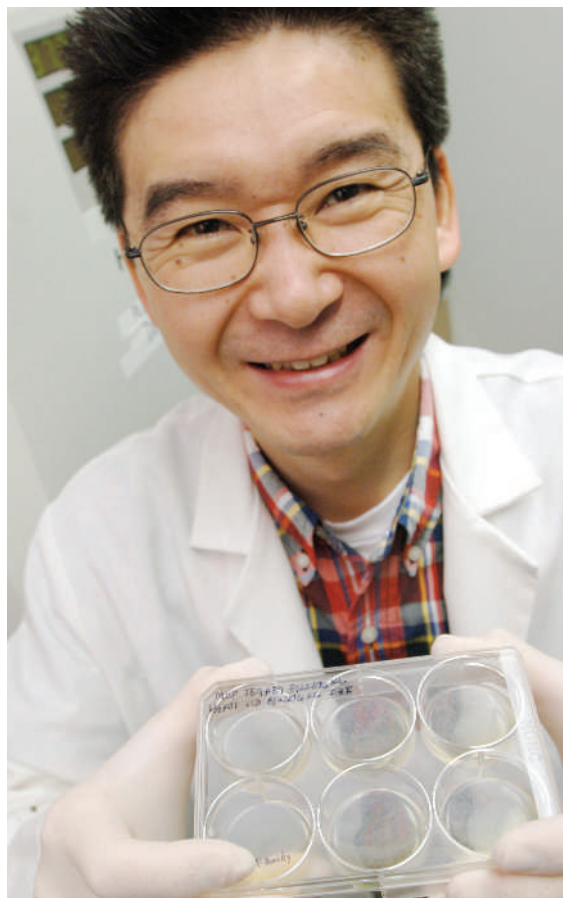
McNerney is running on transforming the district — part of which is making a transition from farming to commuterland — into a hotbed of alternative energy. "We can develop the Silicon Valley of new-energy technology right here," he says. The area already has an established wind-power industry in the shape of Altamont, one of the first wind farms in the United States; McNerney would like to expand this sector, as well as pursue biofuels.

He says that environmentalist alarm at Pombo is giving him a real boost. Californians, after all, are famous for being ahead of the green curve in the United States. "People here are more aware of environmental issues," says McNerney. "It is easy to appreciate nature in a state like California, where we have good weather a lot. And there is a fragility that we can see."

McNerney may not be a career politician, but he has enough campaign savvy not to emphasize his PhD. "I think that it carries a little bit of a disadvantage," he says. "They look at a maths professor as someone who is kind of overbearing and scary." **E.M.**

J. A. FINLEY/AP (MCCASKILL)

L.G. PATTERSON/AP



Stem-cell research is popular with many voters...

N. LANE/AP



...and Kerry Healey hopes to profit from that popularity in Massachusetts.

just a Democratic front, its leaders insist the group is non-partisan. "There have been very strong advocates for science from both parties," says founding member Peter Agre, a Nobel-prizewinning chemist and vice-chancellor for science and technology at Duke University Medical Center in Durham, North Carolina. "This is not a Republican issue or a Democratic issue. It's an American issue."

John Marburger, President Bush's science adviser, agrees — and argues that the Bush administration has done a fine job of advancing various scientific initiatives. "Science has wide bipartisan support in this country," he says. "Scientists benefit from that." He points to the 'competitiveness initiative', meant to keep the United States at the forefront of science and technology, which is currently winding its way through the Republican-led Congress. If brought into effect, it would mean an 8% increase for the National Science Foundation's budget and a 15% boost to funds for the Department of Energy's Office of Science in 2007 (see *Nature* 439, 644–645; 2006).

But if science as an overriding issue is hard pushed to gain traction, specific issues — in particular, stem cells and energy — could play an important role in some tight races. In a neck-and-neck national election, with the Democrats enjoying a lead in public opinion but requiring 15 seats to take control of the House and 6 seats to take control of the Senate, it's possible that the tactical use of these issues in specific races could make a difference.

Culture shock

Meanwhile, in various states, scientific issues are turning up on important ballot initiatives. In Missouri, a high-profile measure would amend the state constitution to protect stem-cell research (see 'Running on science', previous page). In California, leading researchers are pushing for a clean-energy ballot initiative that would put more than \$1 billion towards research on alternative energy sources (see 'Good times for green energy', opposite). These initiatives do not necessarily lend themselves to the big-picture analysis of a 'war on science', but they should reveal how Americans are thinking about various ways in which research can have a practical impact on their lives. And that could, perhaps, influence candidates as they begin strategizing for the presidential race of 2008.

More so than any other line of research, stem cells have brought candidates from both parties to the front lines of science. Wilson is far from the only Republican now distancing herself from Bush on the issue. "You won't see any Republicans in competitive races touting the fact that they supported Bush on human embryonic stem-cell research," says Sean Tipton, president of the Coalition for the Advancement of Medical Research, a Washington-based advocacy group that supports the research. "A lot of candidates recognize that it plays into real weaknesses for Bush and his party."

In several governors' races, Republican candidates are reaching the same conclusion, sometimes moving to support generous state funding for stem-cell research. In Maryland, governor Robert Ehrlich stood quietly by in April 2005 as a bill providing \$25 million in research funds for such research died in the face of conservative opposition in the state legislature. But this January, he came out vocally in support of the work, and in April he signed into law a bill providing \$15 million in state funding.

Good times for green energy

California is again trying to leapfrog the rest of the United States with a dramatic new research initiative. Last time, in 2004, it was stem cells. This time it is alternative energy.

On 7 November, California voters will be asked to approve a proposition that would raise taxes on the oil industry in order to create a fund to reduce dependence on fossil fuels. Of \$4 billion drawn from the fund during the next ten years, about \$1.1 billion would be available for research on renewable or efficient energy technologies.

Proposition 87 mirrors aspects of the 2004 California measure when voters approved \$3 billion for stem-cell research. In both instances, the state has moved when the federal government balked at aggressively funding research.

"California has a clear mission," says Daniel Kammen, director of the Renewable and Appropriate Energy Laboratory at the University of California, Berkeley, and the first of many scientists recruited to the proposition's cause. "The nation will be looking to emulate the California model."

Under Proposition 87, a 'severance tax' would be levied on the 200 million barrels of oil annually produced in California. California is currently the only state without such a tax, although its oil companies are taxed in other ways. The severance tax could be levied for a decade, but given the recent high price of oil, proponents anticipate the \$4 billion would be raised in less than six years. The money would be divided between two main programmes: the \$1.1-billion research allocation, and \$2.3 billion for creating businesses that would reduce use

of fossil fuels by 25% by 2017.

A governing board, similar to the state's stem-cell oversight group, would direct exactly where the money goes. There would be places on the nine-person board for, among others, a research scientist with expertise in the area, a business academic, a public-health expert and a venture capitalist.

The measure is part of a nationwide mosaic of fledgling state laws designed to cut fossil-fuel use and offer viable renewable energy sources. Already, California forces utility power-plant operators to cut emissions if they sell electricity into the state, no matter where the generators are located; similar programmes are under way or being planned in places such as New England, Arizona and New Mexico. Also in California, the country's first law to reduce greenhouse-gas emissions will come into effect on 1 January 2007.

The Proposition 87 campaign is coordinated by Anthony Rubenstein, an erstwhile Hollywood screenwriter who previously ran an urban renewal organization in Los Angeles. It has been generously funded by Steve Bing, a movie producer, who has pledged an unprecedented \$40 million to the cause. Other notable supporters include venture capitalists Vinod Khosla, a biofuels evangelist, and John Doerr. The oil industry, which claims disingenuously that the proposition would increase prices and more reasonably that it would lead to California importing oil rather than producing it, has put up a similar amount of money in opposition. The campaign seems certain to be the most expensive ever fought on a ballot proposal.

In Massachusetts, lieutenant governor Kerry Healey (Republican) has been at pains to show her support for stem-cell research in her bid to become governor. In August, she broke publicly with the current governor when his administration announced a restrictive interpretation of a 2005 state law governing stem-cell research. A spokesman for Healey called the rules a "mistake" that "could have a chilling effect on those individuals at the forefront of this emerging field".

In some states, the issue is economic as well as political. Jim Doyle, the incumbent Democratic governor of Wisconsin, has pointed at his challenger's congressional vote against expanded federal funding for stem-cell work. Doyle has also emphasized his own role in bolstering research in the state where human embryonic stem cells were first isolated. That's a canny strategy in a state that is "terrified that it's losing jobs and people, and needs biotechnology", says Arthur Caplan, director of bioethics at the University



California's only public alternative-fuels station in San Diego. More to come?

The proposition itself was largely written by leading California energy scientists. After receiving a cold-call at home on Christmas morning in 2004 from Rubenstein, Kammen set about enlisting an impressive array of California's Nobel laureates and other leading scientists in support of the measure. Supporters imagine the proposition's success bringing even more key scientists to the state. "Just like with stem cells, it will lead to top researchers moving here," says chemist Nathan Lewis of the California Institute of Technology in Pasadena.

Three weeks before the election,

polls suggest that the electorate is fairly evenly split; the 20-point lead that the proposition enjoyed in the summer, before the campaign really got under way, has been whittled down. So far, California governor Arnold Schwarzenegger, who is an advocate of clean-energy research and measures to curb climate change, is opposing the proposition, so as not to increase taxes. But those backing the idea are optimistic he may join the bandwagon if polls reflect support — just as he did with the stem-cell initiative, which he backed at the last minute.

Rex Dalton

of Pennsylvania, who has advised several Democratic and Republican campaigns on stem cells. "It's an economic argument dressed up in a stem-cell costume."

In other states, though, Republicans are using opposition to stem-cell research as a traditional pro-life values issue. In the Missouri senate race (see 'Running on science') Republican incumbent Jim Talent hopes that opposition to a stem-cell measure on the state ballot will increase turnout among the conservative voters he needs.

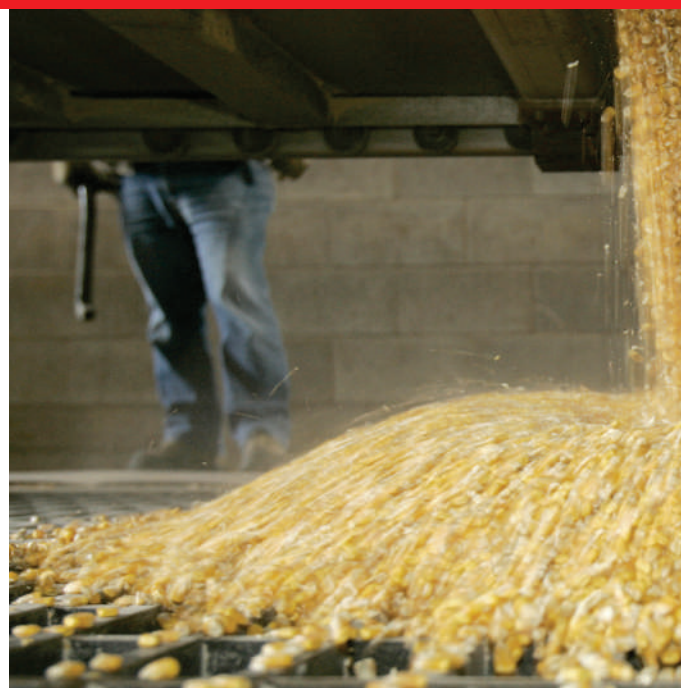
Politicians know that stem-cell issues appeal to many voters. So does the promise of 'energy independence', a catch-phrase that promises no more reliance on foreign oil. Voters have been put off by the ongoing war in Iraq and unrest in the Middle East, as well as record-high oil and gas prices this summer, although these have eased somewhat in recent months. "Usually, we are fighting to get issues up in the forefront," says Massaro. "This year, everyone is thinking

and talking about energy.” Many gubernatorial candidates, including Wisconsin’s Doyle and Ted Kulongoski, Democratic governor of Oregon, are advancing some version of the ‘25 by 25’ pledge — the broad-based push to produce 25% of the country’s energy from renewable sources by 2025.

The issue has come into play in closely fought Senate races as well — and again on both sides of the partisan divide. In Washington state, Democratic senator Maria Cantwell is posing with wind turbines even as her Republican opponent proclaims his support for heavy investment in alternative energy. In Tennessee, Democratic candidate Harold Ford runs adverts wherein he strides across fields of soya beans grown for biofuel. In New Jersey, Republican challenger Tom Kean says that, “unlike President Bush”, he doesn’t think America can “drill its way to energy independence”.

The interest in energy issues runs deep. Earlier this year, the liberal citizens’ group MoveOn.org staged more than 1,000 house parties, asking attendees to name the issues they thought the group should press hard on for the elections. “There were just two issues that came up at every one of those house parties,” says Eli Pariser, executive director of MoveOn’s political action committee. “One was health care and one was energy.” For Pariser, the issue is about more than oil prices and geopolitics: “There is this sense of a grand scientific exploration in the style of the campaign to put a man on the Moon. People are

The push towards using clean energy sources, such as biofuel derived from crops, has been used by some candidates to woo floating voters.



hungry right now to be asked to be part of a big project.”

But some routes towards energy independence involve extracting non-renewable energy sources — such as drilling in the Arctic National Wildlife Refuge, a plan pushed heavily by Ted Stevens, a Republican senator from Alaska. The drive for energy independence shouldn’t eclipse the message of preparing for climate change, argues Alden Meyer, director of strategy and policy for the Union of Concerned

Q&A



Representative Rush Holt is a rare thing in the US Congress — a bona fide scientist building a promising political career. Since his election for the 12th district of New Jersey — the one containing Princeton —

eight years ago, this former physicist and son of a West Virginia senator has garnered several powerful committee slots. Holt has emerged as one of the Democratic Party’s most prominent spokesmen on science, education and security. **Colin Macilwain** asked him about the life of a scientist on Capitol Hill, and what the mid-term elections could mean for science and education.

What difference would it make to science, or to scientists in America, if the Democrats took control of the House of Representatives?

The atmosphere in Washington is more politically partisan than I have seen in half a century, and it even affects things like science. I’ve never believed that science is completely removed from policy or politics. But many scientists would say they are appalled at the way a political game has been made of science, such as intelligent design in the schools, where both the president and some in Congress have said that both this and evolution should be taught. And climate change — until very recently it’s been difficult to get anyone to acknowledge that there is climate change and that there is any connection with human activity.

Are scientific issues arising as issues in campaigns around the country?

Not as major issues, but in my district there is a kind of frustration that we’ve been unable to deal with energy problems

— it might be high fuel prices, but somewhere in the voters’ minds it is connected with a failure to find alternatives to fossil fuels, and a failure to listen to scientific analysis on climate change and that sort of thing.

Do the Democrats have a programme for science, technology and education — and, if so, what is it?

It may not be as well known or as well understood as we would wish. We do have a good message [‘The Innovation Agenda’] released six or eight months ago. It calls for nationwide broadband, a greater investment in research and greatly increasing the number of trained science teachers in the schools.

But isn’t it true that, historically, Republicans are likely to spend more money on research and development?

The president has acknowledged that the physical sciences have languished, but in the latest budget, not much has come through. So I’m not willing to elevate him to the hall of fame.



Scientists in Washington DC. “Unless you work in the global-warming message as well, there are some proposals — such as turning coal into liquid fuel — that could wreak havoc with the environment,” he says. “You have to bring in the longer-term fossil-fuel dependence as well.”

Climate change offers less political mileage than energy independence. That may reflect the current American view: in June, a poll conducted on a number of issues by the Pew

Research Center for the People and the Press found that, although 64% of respondents thought energy policy was “very important” to them, only 44% said the same of global warming. Nevertheless, in tight governors’ races in Rhode Island and Massachusetts, candidates have divergent stances on the Regional Greenhouse Gas Initiative, a seven-state scheme for limiting greenhouse-gas emissions.

A. MANIS/AP

Capitol gains

Back in Washington DC, where several powerful representatives have notoriously sceptical views on climate change, the elections could significantly shift the balance of who gets listened to the most. If the Democrats take back either house of Congress, the chairmanship of all committees will switch from Republicans to Democrats. And chairmen and chairwomen have the power to call hearings on topics of particular interest — or to call witnesses such as novelist Michael Crichton to criticize the current state of climate-change research, as happened last year in the Senate’s Committee on Environment and Public Works.

If Republicans lose control of the House, accusations of scientific politicization could gain a higher profile. “I think there would be more investigations if the House changes,” says Kurt Gottfried, president of the Union of Concerned Scientists in Cambridge, Massachusetts. For instance, California Representative Henry Waxman — a Democrat who has been active in pursuing conflict-of-interest issues at the National Institutes of Health and other agencies — is in line to gain the chairmanship of the House Committee on Government Reform.

Representative Bart Gordon (Democrat, Tennessee), meanwhile, is in line to gain control of the House Committee on Science if the House switches majority. As such, he might call hearings on the accusations of scientific censorship at NASA and the National Oceanic and Atmospheric Administration, says Joel Widder, a policy adviser at the lobby group Lewis-Burke Associates in Washington DC. “The politicization of science and politics affecting scientific decision-making would clearly be issues that he would explore,” Widder says. If the Republicans maintain the majority, the same committee might be headed by former Democrat Ralph Hall of Texas, global-warming sceptic Dana Rohrabacher of California or physicist Vernon Ehlers of Michigan.

Which party wins may also influence how science budgets are distributed among agencies and across disciplines, as Congress is in charge of doling out money for scientific research. But the total pot of money for science isn’t likely to grow, as the United States continues to struggle to pay for the war in Iraq and unexpected expenses such as Hurricane Katrina, on top of a growing deficit. “It’s not like the Democrats are going to open the treasury and fix all the budget problems that all the science agencies are screaming about,” says Widder. “I think that the budget environment is likely to be so constrained that it doesn’t matter who’s in charge.”

No matter what happens on 7 November, the face of US science is likely to change. And on 8 November, campaigners from both parties will be picking themselves up, preparing for the new Congress to convene in January — and realizing it’s never too early to start planning for 2008. ■

Reported by Geoff Brumfiel, Meredith Wadman, Emma Marris and Heidi Ledford.

See Editorial, page 724.

What about involving the public in decision-making? A lot of discussion happens in Europe but it doesn’t seem to get much traction in the United States.

I think it is fair to say, and unfortunate to note, that the public is not driving the science agenda. I wish they were. In the United States we have found ourselves in a position where the public says, ‘science is for the scientists, but not for me’. Not often do any non-science or non-engineering constituents come to me with science or technology on their list, and I imagine that’s true for other members of Congress as well.

You’ve said that most people in Congress tend to view science as a special interest, albeit an intelligent one. Have you seen much change in how Congress views science?

The public’s appreciation of science is no better, and maybe a little worse, than a decade ago. In official Washington, scientific subjects have become really politicized. There should be debate about the policy that is derived from science. But, historically, if science puts limits on the choices that are possible, the politicians would accept that. Now, by treating science as just another topic to be dealt with ideologically, or to be part of political

trades, they will even ignore the laws of science.

You decided not to seek a berth on the science committee but to look to more general committees instead. Was that a good choice?

The greatest need here is for scientific expertise in those areas that are not obviously scientific. On funding for NASA or for Antarctic research, we get pretty good scientific advice. But on ‘how do we get reliable elections’ or ‘what is the effect of good transportation planning’, which are to most Americans not obviously scientific, we have the greatest need.

Do you find your fellow congressmen receptive to a bit of scientific knowledge?

Yes. People will listen to me on some subjects more than just an average colleague. Am I as influential as I’d like to be? No — but I work at it.

What is your proudest achievement in the Congress?

It has nothing to do with science, and it is not even easy to describe. But it is building a sense of respect for government, or to put it another way, beating back the cynicism about government, at least within my own district.

A life online

Darwin is the latest eminent scientist to get an online archive. How do these undertakings change our understanding of history, asks **Henry Nicholls**.

When exactly did Charles Darwin first use the iconic and ideology-laden phrase “survival of the fittest”? On page six of the first issue of the first volume of the first edition of *The Variation of Animals and Plants Under Domestication*, published in 1868. The phrase only makes it into *On the Origin of Species* in that book's fifth edition, published the following year.

If you are James Moore, or Janet Browne, or Adrian Desmond, or David Quammen, or any of the other people who have written biographies of Darwin and analyses of his thought, you probably already knew that. Now, though, you don't need years of scholarship to dig out as many such intriguing nuggets as you can imagine. With the launch on 19 October of Darwin Online¹, answering this kind of question becomes a doddle.

The service makes the complete works of Darwin — in the form of both scanned-in pages and searchable text — freely available to anyone with an Internet connection. It also contains a handful of his key unpublished manuscripts, such as the jottings he made during the voyage on *The Beagle*. This kind of resource, allowing in-depth interrogation of a scientist's entire body of work, is becoming increasingly common. Its proponents believe that it will do more than just cut down on bike rides, train trips or even flights to visit physical archives; they expect it to inject new life into historical scholarship.

The ambitious idea for Darwin Online came to John van Wyhe, a historian of science at the University of Cambridge, UK, back in 2002. Although there were plenty of Darwin's works on the web, he recalls, they were spread across many sites with no obvious editorial standards. “It was just utter chaos.” Given the many other great men of science (for they are almost exclusively men) who boast digital archives of their work and correspondence, this was a gap that needed to be filled (see ‘Best of the rest’, opposite). “It just didn't seem right that other figures should have these rich databases and not Darwin,” says van Wyhe.

Van Wyhe aims to add every edition and translation of Darwin's published work to the website by 2009, the bicentenary of his birth and the 150th anniversary of the publication of *On the Origin*. The ultimate plan is to scan in and transcribe a mountain of some 35,000 Darwin manuscripts dispersed around the world, including drafts of *On the Origin*, descriptions of his experiments and even his domestic

accounts. The aim is to host everything Darwin ever wrote except for his private letters, which the long-running Darwin Correspondence Project, also in Cambridge, is dealing with on its own website².

By the standards of scholarship in the humanities, as opposed to those of the well-equipped lab, these projects do not come cheap. Van Wyhe has a grant of £286,000 (US\$532,000) from Britain's Arts and Humanities Research Council (AHRC), although he will need more to finish off the project as he wants it. The Newton Project, which aims to be the definitive, open-access repository for Isaac Newton's writing and correspondence, has received grants from the AHRC that total £900,000. This funding should be enough to transcribe, footnote and publish online his vast collection of theological writings (almost half of what he wrote) and more than half of his unpublished optical writings by 2010. As with Darwin Online, the long-term vision is to publish everything Newton ever wrote — although coding up the mathematics will be a particular challenge, notes Robert Iliffe, the historian from Imperial College London who acts as the project's editorial director.

Iliffe is adamant that the venture offers serious value for money. In the past, thousands of pounds would be thrown at producing a print edition of Newton's work, with only a few institutions prepared to stump up the cash to buy a copy. “The online medium brings with it such a huge audience,” he says.

Heavy hitters

That's certainly the experience at the Einstein Archives Online³. This site currently boasts more than 3,000 scanned-in pages from almost 1,000 of Einstein's scientific and non-scientific works. Even though he wrote almost exclusively in German, there is huge interest in these documents, says Diana Buchwald, director of the Einstein Papers Project at the California Institute of Technology in Pasadena. When the Einstein Archives Online went live back in 2003, there were nearly 250,000 unique user sessions in the first five days, she says.

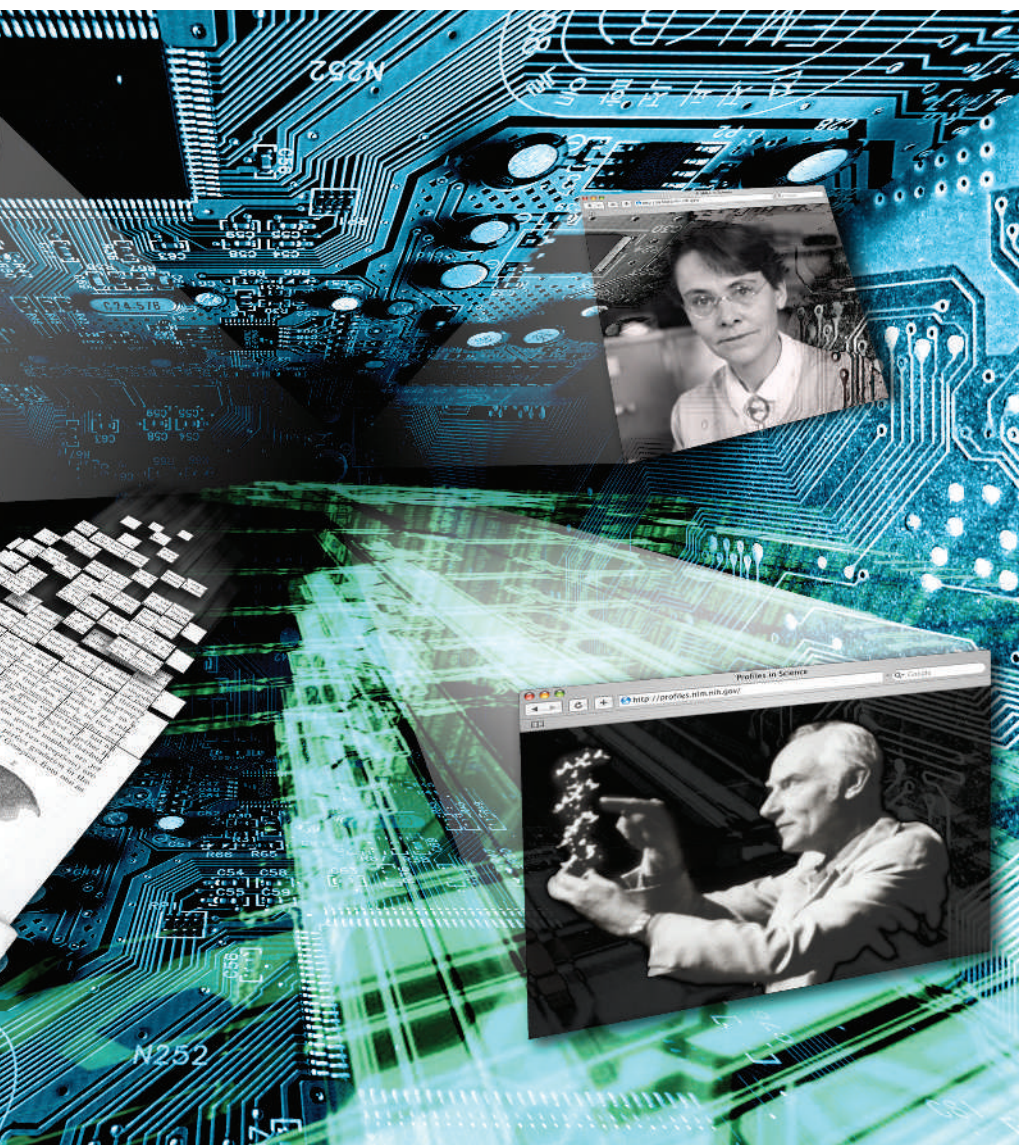
Darwin Online is ready for a similar surge of activity as both supporters and distorters of Darwin's ideas make use of the ability to search through the published material. There are several items that should be of particular interest to Darwin fans, notably the previously unpublished *aides-memoires* he made as *The Beagle* explored the Galapagos Islands. The original notebook has been lost, probably



stolen in 1983, but the University of Cambridge's library has a microfilm of it. It is also possible to inspect all his published illustrations; the visually impaired can download audio files of all the texts.

The creationist faithful would do well to take a look, says van Wyhe. “If people feel so strongly about Darwin, they should actually take the time to read his own words rather than relying only on the interpretations of others.” Even if this doesn't convert them to evolution by natural selection, it should expose the popular misconception that Darwin had an anti-Christian agenda, he says. “This was not what he was about,” says van Wyhe. “He was simply a scientist trying to explain how the world works.”

The same cannot be said for the defiantly irreligious Francis Crick who, enraged by the decision of Churchill College, Cambridge, to build a chapel, wrote a letter to the college's namesake Winston enclosing £10 towards the building of a brothel to go with it. Unfortunately, this letter is in the Churchill archive rather than the Crick archive now being catalogued at the Wellcome Library in London. As



W. FERNANDES, P. JACKMAN/MAYANG.COM

Contemporary history raises other challenges, most notably that of copyright. In many instances, it's simply not clear who owns it. "You may own the physical letter, but the person that wrote it owns the copyright," says Peter Hirtle, intellectual-property officer for Cornell University Library in Ithaca, New York. In an archive with many correspondences, "you have potentially hundreds and hundreds of copyright owners". Consequently, publishing outgoing and not incoming letters is the only practical option for many correspondence projects.

But even if near-complete, searchable archives are to remain the privilege of the long dead, they still offer new scholarly possibilities. Last year, Albert-László Barabási, professor of physics at Harvard University's Dana-Faber Cancer Institute in Boston, and a colleague used data from the Darwin Correspondence Project and Einstein Papers Project to compare how long these men took to reply to incoming letters⁵. The ability to use the dates of documents to map a scientist's writing and correspondence patterns might give new insights into how he organized his life and work, says Barabási. As van Wyhe will be able to tell something about who is reading which of Darwin's texts online, research into the researchers will be possible, too, through studies of the way the resource is used. Will everyone look for survival of the fittest, or will people dig into ever more intriguing byways? What grandeur may be found in this new view of a life?

Henry Nicholls is a freelance science writer in London.

1. <http://darwin-online.org.uk>
2. www.lib.cam.ac.uk/Departments/Darwin/index.html
3. www.alberteinstein.info
4. Ridley, M. *Francis Crick: Discoverer of the Genetic Code* (HarperPress, 2006).
5. Oliveira, J. G. & Barabási, A.-L. *Nature* **437**, 1251 (2005).

yet, only a sliver of the Crick archive is accessible online, by means of the website of the US National Library of Medicine. "It's a fantastic new tool not available to previous generations of researchers," says Matt Ridley, who has just published a biography of Crick⁴. "I longed for even more correspondence to be there."

Contemporary conundrums

But Crick and other contemporary giants raise archival problems that Darwin and Newton avoid. "Indexing becomes more difficult as the corpus gets larger," says Darwin Stapleton, executive director of the Rockefeller Archive Center in New York, whose father named him after the great nineteenth-century naturalist. The archive of a contemporary scientist such as Crick's colleague James Watson may contain more than a million items — including e-mails, spreadsheets, slide presentations, audio recordings and TV appearances, all of which present their own indexing challenges. "What can be done with the Darwin collection cannot be done with a twentieth-century figure," says Stapleton.

Best of the rest

In addition to the online archives for big hitters such as Darwin, Einstein and Newton, there are several other projects intent on revealing the writings of famous scientists.

The Robert Boyle Project (www.bbk.ac.uk/boyle) hosts about a fifth of Boyle's writings, while his 'work diaries' with annotated transcriptions benefit from virtual page-turning wizardry at www.livesandletters.ac.uk/wd.

Jean-Baptiste Lamarck's archive (www.lamarck.cnrs.fr) contains digital photographs of the great biologist's 19,000-strong herbarium collection as well as scanned-in pages from his manuscripts and transcriptions of much of his published work.

The Panopticon Lavoisier (<http://moro.imss.fi.it/lavoisier>) presents much of Antoine-Laurent Lavoisier's published and unpublished

chemical work, hundreds of photographs of his laboratory instruments and mineral collection, and a catalogue of his own library.

The Thomas A. Edison Papers Project (<http://edison.rutgers.edu>) holds scans of 180,000 digital images from the 5 million pages contained in the Edison archive.

The collection of Eva Helen and Linus Pauling papers (<http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/index.html>) contains some 500,000 items, of which around 100,000 have been digitized. Not all of these are available online.

The US National Library of Medicine (<http://profiles.nlm.nih.gov>) has a repository for a selection of works from twentieth-century leaders in biomedical research and public health, including Francis Crick, Oswald Avery, Joshua Lederberg and Barbara McClintock.

Nature: the many benefits of ecosystem services

SIR — In his Commentary “Selling out on nature” (*Nature* **443**, 27–28; 2006), Douglas J. McCauley dismisses the importance of ecosystem services as a tool in conservation and resource management. The author correctly notes that market-based approaches to conservation are no panacea, as has also been concluded by the Millennium Ecosystem Assessment (*Ecosystems and Human Well-Being: Synthesis*, Island Press, Washington DC, 2005). But he goes on to conclude that there is no value in factoring ecosystem services into decision-making, and that indeed they represent a harmful diversion from a more traditional focus on the intrinsic and aesthetic values of nature. We, the assessment panel of the Millennium Ecosystem Assessment, believe that these conclusions result from three errors in reasoning.

First, McCauley assumes that conservation arguments based on ecosystem services are cast only in economic terms. In practice, although it is possible to calculate the economic values of some ecosystem services, this can't be done for others, including many of the cultural services provided by ecosystems. Proponents of ecosystem services argue that it is folly to ignore real economic costs and benefits of decisions. Deliberative decision-making processes are necessary to allow economic, cultural and intrinsic values to be weighed.

Second, McCauley assumes that conservation efforts based on ecosystem services rely only on market-based approaches and hence are always subject to the vagaries of the market. This is not the case. The useful roles played by a watershed in water purification, a woodland for recreation or a forest for carbon sequestration are just some of the many factors used to help convince a government of the merits of protecting certain areas from development. For example, although it would be possible to argue that the coastal wetlands of Louisiana should be protected for their intrinsic value, it is logical — and probably far more effective — to add the utilitarian argument that those wetlands also provide a valuable service in protecting coastal development from storms.

Finally, McCauley assumes that the growing interest in ecosystem services is relevant only to the goal of biodiversity conservation. In practice, scientists, managers and decision-makers are increasingly using the concept of ecosystem services because of its broad usefulness across a wide range of resource-management issues, not just biodiversity protection.

For too long, scientists and managers have tended to view the world as either protected because of the intrinsic or aesthetic value

of the area, or developed for its utilitarian benefits. The reality, of course, is that our planet is a mosaic of systems providing people with different bundles of ecosystem services and disservices. We cannot manage these systems effectively if we do not actively seek to measure the flows of these services, examine who is benefiting from them, and consider a range of policies, incentives, technologies and regulations that could encourage better management and sharing of the benefits.

Historically, conservation has largely relied on the considerations of intrinsic value that McCauley sees as the only solution. This has been manifestly insufficient as a response to the increasing threats to biodiversity, particularly in the world's poorest regions, where considerations of intrinsic and spiritual values are often trumped by the needs for survival or used to exclude significant segments of the population from the benefits from their ecosystem resources. It is time to add to the mix other approaches based on a fuller consideration of ecosystem services and options for distributing costs and benefits that may result.

Further information, and details of the signatories, are available at www.maweb.org/en/about.people.panel.aspx.

Walter V. Reid

Conservation and Science Program, The David and Lucile Packard Foundation, 300 Second Street, Los Altos, California 94022, USA

This letter was also signed by:

Harold A. Mooney, Doris Capistrano, Stephen R. Carpenter, Kanchan Chopra, Angela Cropper, Partha Dasgupta, Rashid Hassan, Rik Leemans, Robert M. May, Prabhu Pingali, Cristián Samper, Robert Scholes, Robert T. Watson, A. H. Zakri, Zhao Shidong.

Nature: ecosystems without commodifying them

SIR — Douglas J. McCauley, in his Commentary “Selling out on nature” (*Nature* **443**, 27–28; 2006), suggests that love for nature is incompatible with valuing nature in terms of its contributions to human well-being. But there is no such conflict. Nor is valuation of ecosystem services a panacea; rather, such valuation is one piece of helpful information in the complex task of sustainably managing our natural assets.

Valuing ecosystem services is not identical to commodifying them for trade in private markets. Most ecosystem services are public goods (non-rival and non-excludable), which means that privatization and conventional markets work poorly, if at all. Nevertheless, knowing the value of ecosystem services is helpful for their effective management, which in some cases can include economic incentives, such as

those used in Costa Rica's highly successful system of payment for these services (see www.conservation.org/xp/frontlines/partners/03150604.xml).

It is incorrect to suggest that ecosystem-services reasoning ignores basic ecology; on the contrary, it embraces ecology and the co-dependency of humans and other species. It is also incorrect to suggest that conservation based on protecting ecosystem services is betting against human ingenuity. The study of ecosystem services has merely identified the limitations and costs of ‘hard’ engineering solutions to problems that in many cases can be more efficiently solved by natural systems. Pointing out that the ‘horizontal levees’ of coastal marshes are more cost-effective protectors against hurricanes than constructed vertical levees is only using our intelligence and ingenuity, not betting against it.

The ecosystems-services concept makes it abundantly clear that the choice of “the environment versus the economy” is a false choice. If nature contributes significantly to human well-being, then it is a major contributor to the real economy (R. Costanza *et al.* *Nature* **387**, 253–260; 1997), and the choice becomes how to manage all our assets, including our natural and human-made capital, more effectively and sustainably (R. Costanza *et al.* *BioScience* **50**, 149–155; 2000).

I do not agree that more progress will be made by appealing to people's hearts rather than their wallets. Ecosystems are critical to our survival and well-being for many reasons — hearts, minds and wallets included.

Robert Costanza

Gund Institute of Ecological Economics, Rubenstein School of Environment and Natural Resources, The University of Vermont, 617 Main Street, Burlington 05405, USA

Nature: poorest may see it as their economic rival

SIR — The moral imperative of saving species and protecting nature, as put forward by Douglas J. McCauley (“Selling out on nature” *Nature* **443**, 27–28; 2006), must be weighed against the moral imperative of saving people. Typically, it is the poorest members of our world community who are most affected by efforts to protect nature, and who suffer the most when ecosystems are degraded.

The conservation debate cannot be reduced to a choice between protecting nature or making an extra million for a yacht or villa. If it were, then perhaps moral arguments alone would be enough to protect the environment. The reality is that poor people are deforesting vast areas of tropical forest for subsistence agriculture, members of indigenous tribes are killing endangered wildlife and out-of-work

fishermen are converting mangrove forests to shrimp farms. Moreover, biodiversity is greatest in the very areas where human populations are most dense, most rapidly growing and most impoverished (R. P. Cincotta, J. Wisniewski and R. Engelman *Nature* **404**, 990–992; 2000).

McCauley does not acknowledge that economic valuation of ecosystem services can provide the data and tools needed to make human well-being part of the design of conservation projects. Although win–win scenarios are hard to find, it is important that we take the care to quantify ecosystem services, so that those situations in which both humans and biodiversity benefit can be identified and promoted. Moreover, if fundamental economic concepts such as GNP could be reformulated to reflect ecosystem services, then nations might embark on policies that better protect their natural capital assets. The economic valuation of ecosystem services is simply a way of getting everyone's moral imperatives on the same page. It is a way of recognizing that conservation must be accomplished in a just and fair manner, in a way that does not pit the basic needs of humans against nature.

Attention to ecosystem services is not equivalent to venal worship of the dollar. Instead, it provides an entry into market incentives, government policies, better-designed conservation projects and a broader constituency for conservation that reaches beyond the affluent Western world. Conservationists who promote valuation of ecosystem services have no intention of selling out on nature — we just want to make sure it is correctly valued.

Michelle Marvier*, Joy Grant†, Peter Kareiva‡

*Santa Clara University, Santa Clara, California 95053, USA

†The Nature Conservancy, 4245 North Fairfax Drive, Arlington, Virginia 22203-1606, USA

Nature: McCauley replies

In my Commentary “Selling out on nature” (*Nature* **443**, 27–28; 2006) I argue that ecosystem services can and should be cautiously applied in certain contexts to advance nature conservation. To characterize this discussion as a polarized face-off between proponents of ecosystem services and advocates for nature's inherent values is to have misunderstood my viewpoint. I offer below some responses to specific points raised in criticism of my position.

I concede to anyone wishing to argue that the cultural, historical and aesthetic values of nature can in fact be considered “ecosystem services”. This difference seems largely semantic. Call them what you wish, so long as they are made important in conservation.

I cannot agree that the citizens of

developing nations are unable to recognize the inherent worth of nature or act to protect it. Many so-called ‘poor’ cultures have intense legacies of respect for and stewardship of nature. Furthermore, this viewpoint ignores centuries of sacrifice made by severely impoverished people to morally inspired causes such as religion, politics and social movements that did not make them money or directly improve their livelihoods. I simply do not believe that nature is a luxury of the rich.

Although I agree that there is no harm in emphasizing the usefulness of nature, I reassert that there may be harm in overemphasizing this utilitarian worth. The roof of the Sistine Chapel is stunningly beautiful and has much intrinsic value. It also serves to keep the rain out of the church. Pointing out the practical benefits that nature confers will assist conservation so long as these are properly contextualized with and do not harmfully obscure the importance of nature's immense aesthetic worth. Using a diverse approach in conservation will be useful in some circumstances, but in my opinion would not be as necessary if we worked sufficiently hard in the first instance to educate people about nature's intrinsic value.

My point in writing the Commentary was twofold: first, to encourage a critical review of the strengths and weaknesses of ecosystem services; and second, to more properly articulate an appropriate role for ecosystem services in conservation. I thank the authors of these Correspondence letters for assisting with both tasks.

Douglas J. McCauley is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

Melanoma rates remain high in Australia

SIR — Your Editorial “Preventing cancer” (*Nature* **442**, 720; 2006) is surely incorrect when it argues that because of public-health campaigns melanoma has less of a health impact in Australia than in Britain.

As the populations of Australia and Britain differ, it is necessary to quote rates rather than absolute numbers. If this is done, melanoma mortality is two to three times higher, and incidence rates around four times higher, in Australia than in the United Kingdom. The absolute number of deaths is higher in Britain because more people live in Britain than Australia (see B. Armstrong in *Textbook of Melanoma* (eds J. F. Thompson, D. L. Morton and B. B. R. Kroon) 65–80; Dunitz, London, 2004).

Jonathan Rees

Department of Dermatology, Lauriston Building, University of Edinburgh, Lauriston Place, Edinburgh EH3 9HA, UK

Getting the public on board for cancer screening

SIR — Your News Features on cancer (*Nature* **442**, 735–743; 2006) highlight important developments in research since President Richard Nixon declared war on cancer in 1971. Numerous epidemiological as well as case-control studies have confirmed that early intervention translates into better survival.

However, when looking at biomarkers — DNA or proteins that could indicate pathological processes — it is important to distinguish between those that assess ‘risk of cancer’ and those that are suitable for screening. In the former, the individual may already be at risk, harbouring a pre-malignant condition such as adenomatous colonic polyps. Although benign, these polyps have the potential for malignant transformation over time. Screening, on the other hand, aims to detect disease in those who have no symptoms: the smear test, which picks up precancerous changes in the cervical lining, is a successful example.

The public needs to be aware of the differences, otherwise uptake will be low. This is a real problem with breast cancer, for example: uptake of screening is low among women from ethnic minorities, who consequently have higher-than-average rates of breast cancer in both the United Kingdom and the United States (V. N. Thomas *et al. Int. J. Palliat. Nurs.* **562**, 564–571; 2005, and L. Jandorf *et al. Cancer* **107**, 2043–2051; 2006).

Strategies for the prevention of cancer require biomarkers of early precancerous changes within normal tissue. Ideally, these should be linked to environmental factors that can be modified, such as diet. In colorectal cancer, for example — where 75% of variance can be explained by diet — the ideal biomarker should reflect differences in dietary exposure. Most people are interested in food and might be prepared to take the recommended action.

The successful test must not only be sensitive and specific for the disease but also appealing to the public; invasive procedures, for example, are less likely to win widespread acceptance. Lessons can be learnt from breast-cancer screening about improving education and awareness. Rightly or wrongly, the success of a screening biomarker is largely driven by public opinion, which is often poorly informed. Thus, the added task for scientists/clinicians is to convince the public of the efficacy and importance of any new biomarker.

Ramesh P. Arasaradnam

Human Nutrition Research Centre, School of Clinical Medical Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK

COMMENTARY

R. CARR/AP



Presidents Bush (left) and Clinton both set up major science initiatives in their second terms. But there are advantages to being part of the administration's initial agenda.



T. SLOAN/AP/GETTY IMAGES

Planning for US science policy in 2009

To maximize the resources allocated to science and technology during the next US administration the science community must prepare now, argues **Thomas Kalil**.

In less than two-and-a-half years, the next president of the United States will take the oath of office and deliver his or her inaugural address. In early 2009, the president and a small group of sleep-deprived aides will submit a multi-trillion-dollar budget, nominate or appoint senior advisers and members of the cabinet, and establish the administration's initial policy priorities. As a result of my eight years as a science and technology policy adviser to President Clinton, I am convinced that the scientific and technical community should begin to plan now for this transition.

A president can devote additional resources to science and technology at any point during their time in office. For example, President Clinton's National Nanotechnology Initiative (NNI) and President Bush's American Competitiveness Initiative both came in their second terms. However, there are three advantages to being part of an administration's initial agenda. First, presidents often establish a broad framework for their fiscal policy in the pivotal first 100 days of their administration. If science and technology is not part of this framework, it will be harder to add it later. Second, if a president

pledges to increase support for research, this commitment will have greater financial impact if made early. Finally, initiatives do not always survive the transition between administrations. Presidents are often more attracted to new policy directions for which they can claim sole credit.

Attention seeking

Although there is no guarantee that science and technology will be a top priority for the next president, there are prominent voices in both the Democratic and Republican parties that are concerned about US economic competitiveness, and are therefore proponents of increased investment in research and development. This is not the only rationale for supporting science and technology. The quest for new knowledge is a worthy end in itself, and can play a key role in national security, preserving the environment, and allowing Americans to live longer, healthier lives. But

since the end of the Cold War, the economic rationale (the contribution of science and technology to growth, productivity, job creation and international competitiveness) has grown in importance.

How the community prepares for the transition is critical for several reasons. First, the standard message of the research community to political leaders ('Give us 7% more money than you gave us last year') is not very compelling. The president and his senior advisers are forced to make difficult budgetary trade-offs

between policy proposals with concrete goals, such as reducing class size in the early grades, putting 100,000 more police on the street, or improving embassy security. In this environment, arguing that a

given science agency should have a 7% increase in its budget will not grab the attention of the president.

During the Clinton presidency, White House advocates for research were far more

"There are prominent voices in both the Democratic and Republican parties that are concerned about US economic competitiveness."

successful in making the case for concerted efforts such as the NNI than in arguing solely for across-the-board increases. Support for targeted areas such as nanoscale science and engineering and information technology can benefit many disciplines, by driving broader increases in the budgets of key science agencies. And while I believe it is essential for the government to support purely curiosity-driven research, this is more likely to happen in an environment in which politicians, the press and the public understand how use-inspired research can help address pressing national and global challenges.

There are risks and limitations associated with this approach. For example, an administration can be overly prescriptive and choose a technical solution to a particular problem, such as the Bush administration's 2003 Hydrogen Fuel Initiative, which touted hydrogen as a means of reducing US dependence on foreign oil and "the key to a clean energy future". A 2004 National Academy of Sciences report, *The Hydrogen Economy*, concluded that a transition to hydrogen was desirable but "the impacts on oil imports and CO₂ emissions are likely to be minor during the next 25 years". Policy-makers should set the goal (reducing greenhouse-gas emissions, say), and allow the research community to pursue the most promising approaches to meeting that goal.

Second, thoughtful and well-grounded science and technology initiatives that are crafted with significant input from the research community can easily take one to two years to develop. Winning support for a new research initiative usually requires determining the rationale for government support, the existing level of funding, the scientific and technological goals of the initiative, the most promising research directions and multi-year budgets. Policy-makers must also determine the appropriate division of labour between science agencies, and a sensible mix of funding strategies, balancing support for individual investigators with research infrastructure and university-industry partnerships. Unless some of this homework has been done in advance, science and technology initiatives are at a disadvantage during the early days of an administration.

A risky business

Below are some of the steps that the scientific community could take to position itself to shape and inform policy during the next administration. I don't want to suggest that this list is comprehensive, but I hope that it will stimulate discussion. First, the community could identify a number of candidate topics for new or expanded research initiatives. To the extent that this has not already been done, the community should also suggest research agendas for each one. Some of these should be

motivated by economic or societal grand challenges, such as the need to generate terawatts of carbon-neutral energy, rapidly respond to emerging infectious diseases, increase student interest and performance in maths and science, bring the Green Revolution to Africa, spark the next revolution in computing that transcends today's chip technology, and establish the technical foundations for synthetic biology and its applications.

Second, the community should develop detailed recommendations on policy issues that affect science and technology across the board. For example, reports by the Washington DC-based Council on Competitiveness (*Innovate America*) and the National Academy of Sciences (*Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*) have bemoaned the fact that agencies are retreating from support for high-risk, high-return research. The gallows humour among some researchers is that "you have to do the experiment before you write the grant". The research community should identify specific steps that the major science agencies might take to reverse this trend, or how the peer-review system could be modified to encourage riskier proposals. For example, should grant programme managers be given more discretion (or explicit instructions) to allocate some fraction of their total budget to support riskier research? Should programmes such as the National Institutes of Health (NIH) Director's Pioneer Award, which is intended to encourage high-risk research, be adopted by other science agencies?

Finally, despite the likelihood of a tight fiscal environment in 2009, there is at least one opportunity for increased research funding that remains largely overlooked by the science community. The US Department of Defense (DOD) plays a vital role in supporting research in certain critical science and engineering disciplines. The DOD's support for basic research is currently US\$1.5 billion, 0.3% of its total budget. It should be possible for the president to direct the DOD to significantly increase its basic research budget. In recent years, the US Congress has consistently increased funding for DOD research above the president's request, but almost all of this increase has gone to support 'earmarks', in which legislators direct money to specific projects in their home states. Earmarks are rare at agencies such as the NIH or the National Science Foundation.

At a minimum, the president could request an increase in DOD funding for basic research to the same level as is already appropriated

by Congress, and ask for a commitment from Congress that this increase would not be subject to earmarks. With further support for DOD research in the president's budget requests, it should be possible to achieve a doubling in agency funding over a five-year period. In addition, the president could redirect the Pentagon's lead research agency, DARPA, to focus on its original role of funding high-risk, high-return research. Under current management, DARPA has funded university researchers primarily as 'sub-contractors' to defence companies, terminated projects that do not meet predefined milestones in 12–18 months, and increasingly classified the research it does.

How can this be done?

This will require the science community to do its homework. Congress expects a detailed description of any new DOD research proposal costing \$5–10 million or more. To make an effective case for increased support for basic research, scientists should provide the DOD with white papers that answer four questions from the 'Heilmeier catechism': What are you trying to accomplish? How is it done now, and with what limitations? What is truly new in your approach that will remove current limitations and improve performance? If successful, what difference will it make? (George Heilmeier was the director of DARPA in the 1970s and developed a set of questions that is still used today for DARPA programme management.)

An important issue is how the research community could be mobilized to get all of this done. Although it would be inappropriate for the National Academy of Sciences to engage in lobbying, they could take the lead in convening, orchestrating and staffing the agenda-setting exercise I have outlined above. Historically, the academy has responded to specific requests from legislators, so Congress could give the academy additional funding and an explicit mandate. If the academy feels uncomfortable taking the lead on such an endeavour, an *ad hoc* coalition of the major science and engineering societies and presidents of research universities could do the job.

The effort I have outlined here would involve considerable work, and there is no guarantee that it would be embraced by the next president. However, there is a decent chance that it would empower advocates for research and development within the White House and the key science agencies to launch exciting and well-designed science and technology initiatives. The research community and the American people would benefit enormously from such an outcome. ■

Thomas Kalil is special assistant to the chancellor for science and technology at the University of California, Berkeley, and a senior fellow with the Center for American Progress.

"Arguing that a given science agency should have a 7% increase in their budget will not grab the attention of the president."

"Policy-makers should set the goal and allow the research community to pursue the most promising approaches to meeting it."

BOOKS & ARTS

What just ain't so

It is all too easy to underestimate the challenges posed by climate change.

Kicking the Carbon Habit: Global Warming and the Case for Renewable and Nuclear Energy

by William Sweet

Columbia University Press: 2006. 272 pp.
\$27.95

Roger A. Pielke, Jr

One of Al Gore's favourite sayings comes from Mark Twain: "It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." I thought of this when reading journalist William Sweet's *Kicking the Carbon Habit*. The book contains an interesting and thoughtful overview of climate science, but concludes with a fundamentally flawed discussion of climate policy.

Unlike many advocates and scientists, Sweet is respectful of his readers in not relying on false claims of certainty about the state of climate science. Although the introduction includes what now seem to be mandatory (yet misleading) references to Hurricane Katrina and Kilimanjaro, the middle part of the book discusses climate science in greater depth, and explains quite refreshingly and accurately that the "seeming unanimity among scientists is, in truth, somewhat deceptive". He acknowledges that uncertainties work in two directions: future climate change could be much more benign than many currently think, or it could be worse. He makes a convincing case that a thoughtful individual can recognize that there are vast uncertainties in climate science, particularly with respect to how the future will evolve, yet still believe that human influence on climate is a problem worth our attention and action.

The book's discussion of policy is regrettably grounded in a fundamental error that surprisingly was not caught in the review process. Sweet, like many others, focuses his policy discussions on recent work by Princeton University's Steve Pacala and Robert Socolow. To present a conceptually clear description of the challenges of reducing emissions growth in the coming decades, Pacala and Socolow described 'stabilization wedges', each of which represents one-seventh of the accumulated future emissions above today's levels out to 2054. They then used the concept of the stabilization wedge as a measuring stick against which to compare 15 policy options; each one could, in principle, result in emissions reductions equivalent to



M. SYKES/ALAMY

Power shift: tackling climate change will mean reducing our reliance on fossil fuels.

one 'wedge'. Pacala and Socolow argued that successfully displacing seven such wedges over the next half-century would be enough to keep open the possibility of stabilizing carbon dioxide concentrations at a level less than twice the pre-industrial value, or about 550 parts per million (p.p.m.). In 2006, carbon dioxide levels are about 380 p.p.m.

However, instead of interpreting Pacala and Socolow's work as offering a trajectory of future emissions that keeps open the possibility of stabilization below 550 p.p.m., Sweet has (mis)interpreted it to mean that seven wedges are "required to stabilize global carbon dioxide at double their pre-industrial value". He seems to have confused some proposed first steps with an act that can deal with the entire challenge. This confusion leads to a dramatic underestimation of the challenge of stabilizing carbon dioxide concentrations at less than twice their pre-industrial value. Sweet suggests incorrectly that breaking the carbon habit can be achieved by implementing 7 of Pacala and Socolow's 15 proposed stabilization-wedge policies: "If the reader accepts only half what [Pacala and Socolow] propose, the problem of greenhouse gas stabilization can in principle be solved."

For their part, Pacala and Socolow recognize that what they have proposed is only a start,

writing that even after the successful implementation of seven of their wedges by 2054, "fossil fuel emissions must decline substantially". But by how much? According to Pacala and Socolow, by about an additional two-thirds over the subsequent 50 years. *Kicking the Carbon Habit* has thus confused stabilizing emissions with stabilizing concentrations — a common error that may have been encouraged by Pacala and Socolow's potentially misleading terminology of 'stabilization wedges' and "solving the climate problem for the next 50 years". The effects of this confusion lead to a misunderstanding of the practical challenges in stabilizing carbon dioxide levels.

In reality, stabilizing carbon dioxide emissions at current levels, as suggested by Pacala and Socolow, would result in a continued linear increase in atmospheric concentrations because carbon dioxide emissions would still far exceed their rate of removal by the oceans and land. Upon completion of the seemingly herculean task of reducing projected global emissions by more than 50% by 2054, by successfully avoiding seven wedges, we would still face the challenge of reducing the remaining level of emissions by another 64% over the next 50 years. To put the stabilization challenge in stark terms, under Pacala and Socolow's most optimistic assumptions for stabilization

at 550 p.p.m., the world will need to reduce its projected business-as-usual emissions by about 1,000 gigatonnes of carbon over the next century. Seven stabilization wedges worth would achieve 175 gigatonnes, leaving a considerable gap, even if the total business-as-usual emissions have been overestimated by a factor of two or more.

It is perhaps Sweet's underestimation of the magnitude of the challenge that leads him to dismiss the prospects for renewables and carbon sequestration in favour of a focus on reducing the emissions from coal. If stabilization at twice pre-industrial levels is to happen, not only will a focus on coal, renewables and sequestration be needed, but many experts argue that there will need to be a fundamental transformation of the global energy system. Based on its misplaced optimism of a relatively quick and easy fix, *Kicking the Carbon Habit* quickly dismisses such perspectives.

Many believe that if climate change can be dealt with relatively easily, then there would also be little need to adapt to it. This sort of thinking may explain why the issue of adaptation plays no role in the book. Overlooking adaptation in any discussion of climate policy is a sign that the challenge posed by climate change has been fundamentally mischaracterized — not only because the world is already committed to some degree of climate change, but also because adaptation makes sense under any future climate scenario.

In the end, *Kicking the Carbon Habit* is a deeply flawed book with considerable potential to mislead its readers about policies related to climate change. This is a shame because its discussions of climate science are both entertaining and informative. ■

Roger A. Pielke, Jr is at the Center for Science and Technology Policy Research, University of Colorado/CIRES, Boulder, Colorado 80309, USA.

interested in developing phage therapy.

Much of this information will be useful for professionals as well as the general public. For example, I was interested to read of the progress made by Omnilytics (a company based in Salt Lake City) in the use and commercialization of phage for the treatment of bacterial spot disease in tomatoes.

However, like many 'popular' science books, the book often lacks a critical analysis of the information presented. For example, Häusler suggests that the spleen filters most phage from the circulatory system, whereas this task is largely done by the liver. He underplays the role of phage in causing human disease. He notes: "Prior to 1900, diphtheria was the most frequent cause of death for German children." Two pages later he states that "phages are viruses that only attack bacteria". He doesn't mention in this context that diphtheria toxin is encoded in a phage genome, so diphtheria is primarily a phage-associated disease. He also implies that the use of lysozymes to kill bacteria is new, but they have actually been used as food preservatives for decades. And there is no merit to his claim that lysin treatment, unlike antibiotic treatments, does not generate resistant mutants. Under laboratory conditions it might be hard to obtain the mutations needed to confer lysin-resistant phenotypes. But as most resistance comes from acquiring new DNA elements from nature, it is easy to see how the overuse of any lysin or antibacterial agent will select for those resistant elements that exist in nature, as has repeatedly happened before.

A more worrying aspect is Häusler's overuse of anecdotal cases. Although he states in the book's preface that the book "is not a health manual whose purpose is to testify to the efficacy of phages", he refers to specific anecdotal cases in a way that suggests that phage therapy was indeed responsible for the reported clinical improvement. These treatments were often carried out in clinical settings that do not meet

The road to phage therapy

Viruses vs. Superbugs: A Solution to the Antibiotics Crisis?

by Thomas Häusler

Macmillan Science: 2006. 256 pp.

£16.99, \$24.95

Sankar Adhya & Carl Merril

The first evidence for a viral-like agent with antibacterial properties was reported by M. E. Hankin in 1896. Found in the Ganges river in India, it was temperature sensitive, capable of passing through a porcelain filter, and could reduce titres of the bacterium *Vibrio cholerae* in laboratory culture. Hankin suggested that it might help to decrease the incidence of cholera in people using water from the Ganges.

The viral nature of such antibacterial agents became clearer following the observation of agents capable of lysing bacterial cultures by Frederick Twort and Felix d'Herelle in 1915 and 1917, respectively. It was d'Herelle who named these agents bacteriophages and championed their use in treating infectious diseases. He and his colleagues introduced phage therapy throughout the world, with major efforts in India, Egypt, the United States and the Soviet Union. Unfortunately, these clinical applications were initiated before certain microbiological aspects of phage strains, such as their narrow host range, were fully appreciated. In addition, clinical investigations in the 1920s and '30s were based largely on anecdotal evidence — scientific and statistically relevant clinical trials were only introduced in the 1940s, primarily in the United States and Britain. The clinical use of phage therapy without the underlying scientific support needed to assure a high level of clinical effectiveness, combined with the development of antibiotics during the Second World War, served to

marginalize the use of phage therapy in most Western countries.

There have been many reviews of phage therapy and some books, but few provide the historical details that Thomas Häusler introduces in this volume. For example, he mentions reports published in the journal *Der Deutsche Militärarzt* during the Second World War on a phage preparation known as 'polyfagin', which was produced by the German pharmaceutical company Behringwerke as a treatment for dysentery. And he reveals that the research of René Dubos on phage therapy in 1942 was supported by a committee of the US National Research Council that was "assigned the task of keeping the country prepared for war". He also interviewed many of the researchers and business entrepreneurs who have been



Troubled waters? Bathers in the Ganges were thought to be protected from cholera by phage.

STEVE BLOOM IMAGES/ALAMY

the current standards required for the certification of marketed pharmaceuticals. One of the cases, which is referred to numerous times throughout the book, involves an individual with a chronic bone infection who was treated by irrigation with solutions containing phage and by having phage-impregnated material placed in the open wounds. It would be interesting to compare such 'phage therapy' with the use of saline irrigations and sterile drains. Although Häusler acknowledges the need for the careful development of therapeutic agents,

along with the need for adequate control experiments, there is a danger that this book might encourage some individuals with infectious diseases to use phage in such a way.

The development of therapeutic phage will need a commitment to meet all the scientific requirements for current pharmaceutical agents. An anecdotal approach, adopted by some of the investigators described in the book and widely used in clinical medicine in the early twentieth century, may retard rather than stimulate the acceptance of therapeutic

applications. The encouraging results of animal experiments have demonstrated phage's capacity to rescue animals with life-threatening infections. Perhaps phage therapy, if carefully developed, could provide some much-needed antibacterial agents. ■

Sankar Adhya is at the Laboratory of Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892, USA. Carl Merrill is emeritus at the Laboratory of Biochemical Genetics, National Institute of Mental Health, Bethesda, Maryland 20892, USA.

Eeyore goes to Washington

The Beginner's Guide to Winning the Nobel Prize: A Life in Science

by Peter Doherty

Columbia University Press: 2006. 320 pp. \$24.95

Peter Parham

Soon after my appointment as an assistant professor, I was interviewed in Palm Springs for a beginner's grant given by a popular American charity. The interviewer, a bejewelled and scarily tanned doctor, looked me straight in the eyes and barked, "So, for what will you win the Nobel prize?" On hesitating to answer, my interview was doomed. I shortly crawled away, to hear nothing more of the matter. Back then, what I could have used was *The Beginner's Guide to Winning the Nobel Prize*. The book's author, Peter Doherty, knows about these things, because in 1996 he and Rolf Zinkernagel were awarded the Nobel Prize in Physiology or Medicine for their joint work on the recognition of virus-infected tissue by T lymphocytes of the immune system.

For the aspiring young scientist, or a student considering a scientific career, Doherty opens the vault to the world of science, explaining how it works and how to get on. His title is only lightly in jest, for the last chapter gives a common-sense set of 18 guiding principles to scientific success that almost any old hand would agree with. As well as continuing



Winning a Nobel prize has given Peter Doherty a platform to air his views on climate change and the environment.

CORBIS SYGMA

his kind, however, Doherty aims to reach a broader readership: those who learn from the movies that "scientists are always mad, bad or quaint nerds who rattle on about controlling the world".

He is not motivated by simple enthusiasm alone, however: like others in the know, Doherty has become deeply concerned with the way governments manipulate science to suit their political goals, and with the way publics at large, paradoxically, combine a pervasive suspicion of science with an equally

pervasive confidence that technology — the application of science — can alone solve all the world's problems.

Doherty is an Australian, a veterinary surgeon and a faculty member at both St Jude Children's Research Hospital in Memphis, Tennessee, and the University of Melbourne. Once described by the eminent immunologist Philippa Marrack in *The New York Times* as "a bit Eeyore-like" (Eeyore being the indomitable, if lugubrious, donkey from A. A. Milne's *The House at Pooh Corner*), Doherty says that

NEW IN PAPERBACK

Nature: An Economic History

by Geerat J. Vermeij (Princeton University Press, \$19.95, £12.95)

On the Origin of Phyla

by James W. Valentine (University of Chicago Press, £22.50, \$35)

"A magnificent book [about] one of the most significant revolutions in the history of life, the Cambrian explosion." Stefan Bengtson, *Nature* **430**, 506 (2004).

Our Inner Ape: The Best and Worst of Human Nature

by Frans de Waal (Granta, £9.99)

"In this excellent book for the public, Frans de Waal tackles some exasperating misconceptions about the evolution of the social behaviour of apes, particularly humans." Robert Sapolsky, *Nature* **437**, 33–34 (2005).

The Tree: A Natural History of What Trees Are, How They Live, and Why They Matter

by Colin Tudge (Crown Publishers, \$27.95)

Pandora's Baby: How the First Test Tube Babies Sparked the Reproductive Revolution

by Robin Marantz Henig (Cold Spring Harbor Laboratory Press, \$15.95)

The Equation That Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry

by Mario Livio (Simon & Schuster, \$15)

"Mario Livio follows [Galois'] brief existence like a sleuth... and gives a panoramic view of the direct, as well as quite remote, applications of group theory." István Hargittai, *Nature* **437**, 34 (2005).

the Nobel prize changed his life, propelling him "into a new and different world of public advocacy for science". During the past ten years Doherty has enjoyed travelling the world, hobnobbing with politicians, governors, business leaders and authority figures from all walks of life to explain the nature of the scientific enterprise to them. In return, these interactions have provided a political education and awakening for Doherty, who now speaks out on the pressing issues of population control, global warming and environmental degradation. In his book, these themes of explaining science to the non-expert and exploring the place of science in modern society are intertwined. Each chapter is an essay that could stand alone but cleverly connects with the

others through the device of the Nobel prize.

This book, which was first published in Australia in 2005, follows *How to Win the Nobel Prize: An Unexpected Life in Science* (Harvard University Press, 2003) by J. Michael Bishop, who with Harold Varmus received the 1989 Nobel Prize in Physiology or Medicine. Despite their similar titles, size and desire to show science's human face to the public, they are very different books, making both for two good reads and an interesting comparison. Bishop's is more of a personal memoir, in which one learns all about his life and interests outside the laboratory. Contrasting with Doherty, Bishop reflects that the Nobel prize "has not enriched my life by any large measure".

Over the past month, would-be winners of a

Nobel prize have been distracted, twitchy and unable to look their fellow scientists in the eyes. That is now all over — at least until the same time next year, when another selected group will be looking forward to a wintry week in Stockholm and the celebrity status the Nobel prize brings. Meanwhile, a headline in *The New York Times* recently announced: "For Quality TV, Mad Scientist Returns". This referred to the planned re-running of a 14-year-old television programme, *Beakman's World*, which aims to introduce the next generation to the joys of science through the persona of an electric-haired, bulging-eyed mad scientist. ■

Peter Parham is in the Department of Structural Biology, Stanford University, Stanford, California 94305, USA.

Pictures from the edge of darkness

Eight photographers enter the twilight zone.

Colin Martin

"Heavenly shades of night are falling, it's twilight time," goes the song lyric. However, twilight occurs at dawn before sunrise each morning, as well as at dusk after sunset, although its effects then occur in reverse. Its duration differs for 'astronomical', 'nautical' or 'civil' twilight. It varies from roughly 120 minutes for astronomers, who need a very dark sky to see all the stars, to 90 minutes for old-style navigators, who need see only the brightest stars to use their sextants, and 60 minutes for civilians, who need street lights when it is only moderately dark.

In *Twilight: Photography in the Magic Hour*, at London's Victoria and Albert Museum until 17 December, eight contemporary photographers are exhibiting around 50 works that explore the dramatic visual effects observed in the sky during twilight. Many of these effects result from the Sun's position below the horizon, which causes Earth's atmosphere to scatter more of the shorter wavelengths of light (from the blue end of the spectrum) than it does during the hours of daylight, leaving more of the longer wavelengths (towards the red end) to reach our eyes.

How we see colour depends on light levels. Photographers always have to overcome the technical limitations of cameras and film, which cannot accommodate to the low light levels and colours of twilight with the flexibility and subtlety of the human eye and brain. They can use this distortion, however, to heighten the psychological effects of twilight, and artists have long attempted to



Photographs by Bill Henson (above) and Robert Adams reveal the remarkable qualities of twilight.

narrative sequence of twilight scenes.

Other artists use the full potential of colour photography to interpret twilight, including the digital manipulation of images. In his series of untitled photographs (2000–03), Bill Henson populates the peripheries of Australian towns with posed groups of androgynous adolescents, creating an edgier twilight zone. "It is not for nothing that twilight is [Henson's] favourite time of day," says Australian critic Peter Craven, quoted in a catalogue essay, "that time when colour still functions as an agent of definition but when it has lost the power to distract the mind with any excess of the sensuous, when everything has bled and receded."

Colin Martin is a writer based in London.

ROSIN OXLEY GALLERY

FRAENKEL GALLERY/MATTHEW MARKS GALLERY

NEWS & VIEWS

CHEMISTRY

The promise of emptiness

Raul F. Lobo

Zeolites are materials with widespread applications. A newly synthesized example has desirably large pores, as well as the virtue of thermal stability, and shows the value of structure-prediction programs.

A microporous crystalline silicate with exceptionally large pores and great potential as an industrial catalyst is described on page 842 of this issue¹. This zeolite is dubbed ITQ-33, and has been discovered by Avelino Corma and co-workers in Valencia.

Zeolites are crystalline oxide materials with a tetrahedral framework structure that contains cavities in which 'guest' molecules can move relatively freely. Such molecules can rapidly be adsorbed, react and desorb, making zeolites particularly useful for catalytic and separation applications. The framework is composed mainly of silica tetrahedra, but many other elements, such as aluminium, titanium or germanium, can be incorporated into the framework structure. Aluminium is particularly valuable in catalysis — if substituted for silicon, it makes the framework anionic, so requiring the presence of readily available cations for interaction with the adsorbed species; if the cation is a proton, the material becomes a strong, solid acid.

As a group, zeolites are the most important of industrial solid-acid catalysts, with many applications in the petroleum and chemical industries. One of their most useful properties is their ability to select molecules based on size (that is, their molecular-sieving properties). But for some applications, such as cracking hydrocarbons, and specifically hydrocracking, this is in fact a limitation, as the large molecules one may wish to react do not have access to the acid sites located inside the crystals. So the zeolite-synthesis community has been

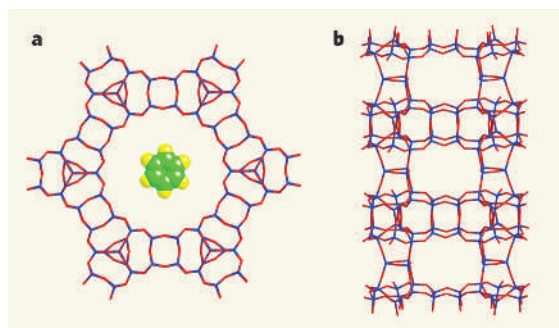


Figure 1 | Structure of ITQ-33, the zeolite synthesized by Corma *et al.*¹. a, View of the framework down the 18-member ring (a pore defined by a ring containing 18 oxygen atoms). The green-and-yellow molecule is benzene, which is 5.5 Å in diameter and is included for size comparison. b, The 10-ring pore windows that interconnect the 18-ring channels.

vigorously seeking materials that, as well as good thermal stability, have larger pores: ITQ-33 is the culmination of decades of progress in this direction (Fig. 1).

In 1988, a related aluminophosphate material with a similar pore size to that of ITQ-33 was reported². But this material, VPI-5, in common with other large-pore, phosphate-based materials prepared since then, has low thermal stability and low acidity. A measure of pore size is how many oxygen atoms form the ring that defines the channel, and several silicate materials possessing pores with 14-member rings (about 8.5 Å in diameter) have been synthesized in the past decade³. A larger example — ECR-34 (18-ring; gallium silicate), with one-dimensional pores about 12.5 Å in diameter — was reported⁴ in 2003.

ITQ-33 has a similar pore dimension to ECR-34, but is much more stable under reaction conditions¹. Moreover, it has medium-size

pores (10-ring, 5.6 Å) that run perpendicular to the 18-ring pores and so allow lateral as well as axial transport of molecules. This is a highly desirable property, because blocked pores can be bypassed using the channels that extend in other directions. The simultaneous presence of 18-ring pores and 10-ring pores perpendicular to them makes ITQ-33 the zeolite with the largest void fraction known (around 0.37 cm³ g⁻¹).

To make ITQ-33, Corma *et al.*¹ had to use all the tricks of the trade that lead to stable microporous materials with large pores⁵. These are the addition of an organic 'structure director' (in this case, hexamethonium) to the synthesis mixture, and the presence of both fluoride ions that stabilize small cage structures in the framework and germanium, an element that the Valencia group and others have shown stabilizes double-four rings (cubes) in the structure of zeolites. But what seems to have

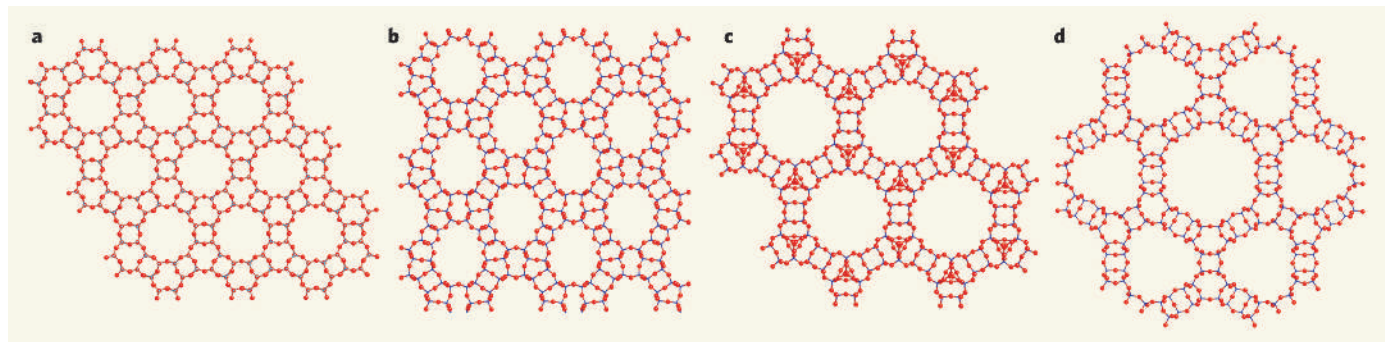


Figure 2 | Zeolite frameworks with pores of different dimensions. a, AlPO-5 with 12-ring pore channels. b, SSZ-53 containing 14-ring pore channels³. c, ITQ-33, viewed along the 18-ring pore channels¹. d, A hypothetical structure (generated by the program of Treacy *et al.*⁷ and available¹⁰ as structure 191_5_277133) containing pores of 18 and 24 rings.

been the key to their success is the use of high-throughput experiments to assess various zeolite compositions: the window of compositions that yielded ITQ-33 is narrow, and outside the common range usually used to prepare zeolites. This demonstration that new materials can be discovered within such a narrow compositional window should lead to the wider use of high-throughput technology in the search for further zeolites.

The role of hexamethonium in the formation of ITQ-33 is intriguing. To date, the general strategy has been to prepare increasingly large organic molecules possessing the rigidity, solubility and stability needed to 'direct' the crystallization of new materials. Typically, the size and shape of the resulting pores corresponds to the size and shape of the organic molecule. ITQ-33, however, is different: hexamethonium is small and flexible, and there is no obvious fit between it and the resulting pore structure. It could be that the hexamethonium molecules pack in such a way as to provide an exact fit for the voids; this is the case, for instance, with VPI-5, which is stabilized by a chain of water molecules that perfectly fit the interior of the pores⁶. Hexamethonium is a simple and relatively inexpensive reagent, and its use bodes well for making ITQ-33 viable for practical application. Other zeolites prepared with organic compounds of similar complexity are used in the petrochemical industry and as additives in catalytic converters.

Another exciting aspect of the latest work¹ is that the structure of ITQ-33 was 'predicted' by algorithms that generate framework structures consistent with the geometrical requirements of a zeolite⁷. In the past year, roughly half of the reported zeolite structures have been previously 'discovered' by these algorithms. It is possible to search the large structural databases generated by these programs for structures with hitherto unavailable properties. An example is given in Figure 2, in which a computer-generated framework with 18- and 24-ring pores is compared with ITQ-33 and other known zeolites. The advent of these powerful algorithms will help in solving the structure of microporous materials, and can make the synthesis of zeolites more 'directed' and perhaps more successful.

Although ITQ-33 has all the characteristics of a good acid catalyst, much work remains to be done to make it practical. The amount of germanium and fluoride required must be minimized or eliminated to reduce manufacturing costs. Better ways of recovering the organic director and recycling it could further increase its potential. Substitution of other atoms in the framework, such as titanium or tin, could expand the range of properties to catalytic reactions such as oxidation and Lewis-acid catalysis.

More generally, ITQ-33 may help us to gain a better understanding of the adsorption processes that occur at the interface between the microporous (pore diameter less than 2 nm)

and mesoporous (pore diameter 2–100 nm) scales. It is at this length scale that the transition between monolayer and multilayer adsorption occurs and where the assumptions of classical adsorption theories can break down. The problem can be approached from the other side, and there are, indeed, mesoporous silicas with ordered and highly uniform pore sizes in the 2-nm range⁸. These materials are, however, difficult to prepare with uniform pores below 2 nm. ITQ-33 bridges these two length scales; and because it is crystalline, and all its pores are — except for defects — identical, one should be able to relate atomic structure precisely to the adsorption isotherms of simple gases. This information could help in the future to interpret adsorption isotherms of other non-crystalline materials that have substantial porosity at the micro-meso transition.

Finally, the discovery of ITQ-33 raises the question of whether we need materials with even larger cavities. Some of the unique properties of zeolites arise from the large curvature of their pores. As the pores get larger, the interaction of adsorbates with the pore walls increasingly resembles the interaction with a flat surface. At some point, the zeolite pore will start to look like the surface of layered aluminosilicates such as clays (albeit without their characteristic hydroxyl groups). Yet

perhaps it is not catalysis or separations where the large-pore materials of the future will find use. Instead, it may be in such niches as sensors or photonics⁹, or where the low-dielectric constant of such materials, arising from their porosity, can be exploited in the manufacture of improved microelectronic devices. The challenge remains to make structures with less and less in them. ■

Raul F. Lobo is at the Center for Catalytic Science and Technology, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716, USA.
e-mail: lobo@udel.edu

1. Corma, A., Díaz-Cabañas, M. J., Jordá, J. L., Martínez, C. & Moliner, M. *Nature* **443**, 842–845 (2006).
2. Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. *Nature* **331**, 698–699 (1988).
3. Burton, A. et al. *Chem. Eur. J.* **9**, 5737–5748 (2003).
4. Strohmaier, K. G. & Vaughan, D. E. W. *J. Am. Chem. Soc.* **125**, 16035–16039 (2003).
5. Corma, A. & Davis, M. E. *ChemPhysChem* **5**, 304–313 (2004).
6. McCusker, L. B., Baerlocher, C., Jahn, E. & Bulow, M. *Zeolites* **11**, 308–313 (1991).
7. Treacy, M. M. J., Rivin, I., Balkovsky, E., Randall, K. H. & Foster, M. D. *Micropor. Mesopor. Mater.* **74**, 121–132 (2004).
8. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
9. Ruiz, A. Z., Li, H. R. & Calzaferri, G. *Angew. Chem. Int. Edn* **45**, 5282–5287 (2006).
10. www.hypotheticalzeolites.net

EVOLUTIONARY BIOLOGY

A kingdom revised

Tom Bruns

An international consortium of researchers has produced an impressive new tree of life for the kingdom Fungi. The results are a testament to cooperation between systematists with different expertise.

On page 818 of this issue, James and colleagues¹ provide a landmark study in fungal evolution. Before now, the only broadly sampled phylogenetic trees of the fungi were based on sequences of a single gene — that encoding the small-subunit (18S) ribosomal RNA. Broad sampling of species is essential, because under-sampling is known to adversely affect the construction of evolutionary trees. However, the quantity and quality of data are equally important, and the 18S data were insufficient to provide strong statistical support for many key branches in the evolutionary trees. In any case, a single-gene tree is always questionable, because different genes can give different views of evolutionary history.

James et al.¹ addressed these problems by collecting sequence data from two additional ribosomal RNA genes, and three protein-coding genetic loci, for a carefully selected sample of 199 species. The results of the combined analyses, outlined in Figure 1, are quite similar to those seen with the earlier 18S data, but

statistical support for some key branches in the tree has improved. This will be a relief to those who have followed the 18S data closely; it means that the new data have produced incremental shifts, not major alterations, in our understanding of fungal evolution.

The fungi, animals and plants are thought to have diverged from each other roughly a billion years ago. They are the only three eukaryotic kingdoms of life that developed multicellularity in terrestrial environments. Like plants and animals, the fungi had to adapt to terrestrial environments from ancestors that were aquatic. But the fossil record for fungi is much the worst; most of them are microscopic with relatively simple morphologies. For these reasons the evolutionary patterns within the fungi were poorly understood before the advent of nucleotide sequence data. It was known that most fungi lacked zoospores, motile cells that are propelled by flagella in water. Therefore the Chytridiomycota, the one aquatic group of fungi that contains flagella, was assumed

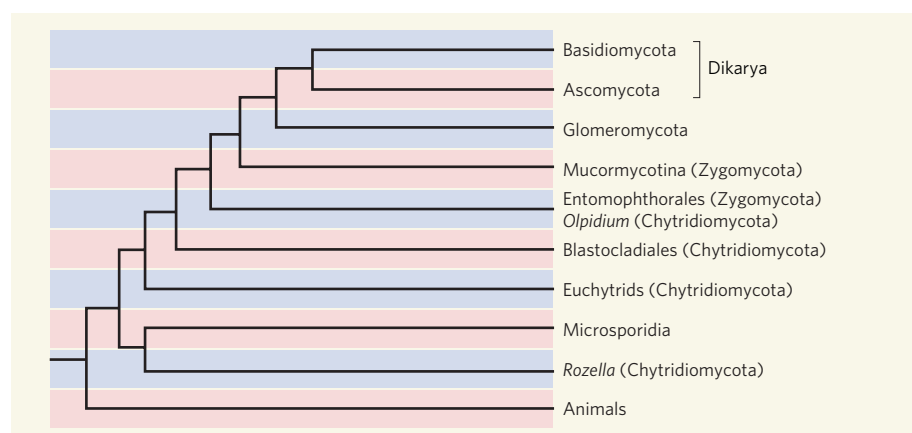


Figure 1 | The main branches of the kingdom Fungi. This highly simplified evolutionary tree shows the traditional phyla — Ascomycota, Basidiomycota, Glomeromycota, Zygomycota and Chytridiomycota. The Ascomycota and Basidiomycota are united as the dikarya, fungi in which part of the life cycle is characterized by cells with paired nuclei. Their closest relatives seem to be the Glomeromycota, a group that was previously included within the Zygomycota. Neither the Zygomycota nor the Chytridiomycota are monophyletic groups; instead they seem to be 'paraphyletic grades' that are grouped only by shared primitive morphologies. Also shown are the microsporidia and *Rozella* branches, which seem to be basal to the all other fungi. (Note that all of these branches are still in need of stronger statistical support. James and colleagues' much more detailed tree¹ appears on page 820.)

to be primitive. This has turned out to be correct but the details of the relationship are complicated.

Both the 18S data and the new multigene analyses show that the Chytridiomycota is paraphyletic — that is, it does not include all the descendants of its most recent common ancestor. But James *et al.*¹ show that a minimum of four independent losses of flagella has occurred; thus one of the key adaptations to the terrestrial environment has actually happened multiple times. Surprisingly, they show that one chytrid, *Olpidium brassicae* (Fig. 1), may lie within the Entomophthorales, a group that includes insect parasites that lack flagella and that is usually considered a subgroup of the Zygomycota. *Basidiobolus*, traditionally a member of the Entomophthorales, had been placed within the Chytridiomycota by 18S data, but is now moved back by the multigene analysis to its more traditional place.

An interesting example of multigene support concerns the placement of the Glomeromycota. These fungi form mutualisms called mycorrhizae with the roots of most plants, and they had been considered to be members of the Zygomycota. The 18S data consistently depicted them as a distinct group closely related to the Ascomycota and Basidiomycota, but there was no statistical support for this placement. The multigene data, however, provide at least bayesian statistical support for the latter relationship (Fig. 1).

The most surprising result concerns *Rozella* (Fig. 2), an obscure genus that is parasitic on other Chytridiomycota. Together with the microsporidia, an enigmatic group of animal parasites, *Rozella* seems to be basal to all other sampled fungi (Fig. 1). There was no reason to expect this, and in this sense the result is reminiscent of the finding by plant systematists that an obscure tropical genus, *Amborella*, is the

sister group to all other flowering plants². These types of result again underline the importance of which species are sampled.

The placement of the microsporidia themselves is another notable result. On the one hand, the analyses with 18S sequences originally put them at the base of the eukaryotic tree, distant from fungi, animals and plants. But this conclusion turned out to be erroneous owing to a confounding factor known as long-branch attraction. Other studies using protein-coding genes had previously placed them in the fungi, but the exact relationship was unclear because of limited sampling within the kingdom^{3,4}. James and colleagues¹ have now improved the sampling dramatically, and show that the microsporidia must be either at the base of the fungal tree, within



Figure 2 | *Rozella allomyces*. This parasite of other members of the same phylum, the Chytridiomycota, seems to be one of the most primitive fungi. Its resting sporangia (spore-producing bodies) are approximately 18 μm across and are shown within a hypha of its host chytrid, *Allomyces*.

the Chytridiomycota, or within the Entomophthorales; in addition they were able to reject eight previously theorized placements within the fungi or outside the kingdom.

There is still room for improvement in two key areas: branch support and taxon sample. Even with six gene loci, many branches remain unsupported or supported only by bayesian statistics, which may give overly optimistic assessments. For many branches it may be possible to increase support by adding additional data, and genomics will be a major contributor. Data from 29 complete fungal genomes were included in the analysis, but this sample is highly biased towards serious pathogens and model genetic systems. With the cost of sequence acquisition dropping, the number of sequenced fungal genomes will increase, and it may be possible to distribute this effort more evenly across the kingdom to provide a better evolutionary sample.

As to the second area for improvement, greater effort needs to be focused on sampling the environment for unknown fungal groups. It is estimated that the kingdom contains 1.5 million species, fewer than 5% of which have been described⁵. If most of the unknown species are members of well-known groups, then the current phylogenetic estimates should be largely unaffected by additional discoveries. However, some entirely new lineages have been recovered by sequence analysis of common but previously unsampled environments^{6,7}: we can't predict how such discoveries will affect our perception of fungal evolution.

The cooperation among researchers that has resulted in the new paper¹ is almost as impressive as the product itself. Systematics can be a fairly balkanized field, with specialists defending their turf or their analytical methods against perceived competitors⁸. However, cooperation has always been common among fungal researchers because the field is woefully underpopulated. The James group included both traditional, morphologically based systematists, who contributed a wealth of knowledge on the organisms, and molecular systematists, who supplied the methodological and analytical techniques. Even Ralph Emerson, who died in 1979, made a notable posthumous contribution: it was his culture of *Rozella*, isolated in 1947, that made the sequence acquisition for this critical branch possible. This fusion of talents was essential to ensure that the broadest possible sample of fungi was selected, and that the data were collected and analysed rigorously. The results represent a proud moment for the field, and will be in the textbooks for some time to come.

Tom Bruns is in the Department of Plant and Microbial Biology, 111 Koshland Hall, University of California, Berkeley, California 94720-3102, USA.
e-mail: pogon@berkeley.edu

1. James, T. Y. *et al.* *Nature* **443**, 818–822 (2006).
2. Qiu, Y.-L. *et al.* *Nature* **402**, 404–407 (1999).

3. Hirt, R. P. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
4. Keeling, P. J. & McFadden, G. I. *Trends Microbiol.* **6**, 19–23 (1998).
5. Hawksworth, D. L. *Mycol. Res.* **105**, 1422–1432 (2001).
6. Schadt, C. W., Martin, A. P., Lipson, D. A. & Schmidt, S. K. *Science* **301**, 1359–1361 (2003).
7. Suh, S. O., McHugh, J. V., Pollock, D. D. & Blackwell, M. *Mycol. Res.* **109**, 261–265 (2005).
8. Hull, D. L. *Science as a Process* (Univ. Chicago Press, 1988).

STRUCTURAL BIOLOGY

Enzyme target to latch on to

Malcolm A. Leissring and Dennis J. Selkoe

Insulin-degrading enzyme is implicated in diabetes and Alzheimer's disease, but few molecular tools exist that can probe its function. A study now reveals its unusual structure and may lead to an expanded toolbox.

Proteases are vital enzymes that have been targeted for the treatment of many diseases. One such protease, insulin-degrading enzyme (IDE), has strong links to diabetes and Alzheimer's disease but has nonetheless proved to be an elusive drug target, despite more than 50 years of intensive research. On page 823 of this issue, Shen and colleagues¹ reveal high-resolution crystal structures of IDE that open the door to the rational design of pharmacological modulators of this protease*. Crucially, the authors show that it might be possible to develop not just inhibitors, but activators as well.

IDE was discovered in 1949 by the physician and biochemist I. Arthur Mirsky². Mirsky reasoned that inhibitors of IDE would be an ideal anti-diabetic therapy, as they would slow the degradation of insulin. In support of this approach, Mirsky found that liver extracts containing an inhibitor of IDE enhance the action of insulin when injected into rabbits³. Thereafter, Mirsky and many others sought to develop potent inhibitors of IDE as potential drugs. Despite these efforts, very few compounds that specifically inhibit IDE are available today, apart from substrates of IDE such as insulin itself. By revealing IDE's active site in unprecedented detail, the crystal structures provided by Shen *et al.*¹ may hold the key to realizing a potent and selective IDE inhibitor.

Recent discoveries, however, raise concerns about the wisdom of inhibiting IDE. Chief among them is the finding that IDE naturally degrades the amyloid- β protein that accumulates abnormally in Alzheimer's disease⁴. Here, it would be desirable to activate rather than inhibit IDE, a strategy that has already proven effective in mouse models of the disease⁵. Moreover, results from different animal models cast doubt on the concept of treating diabetes by chronically inhibiting IDE. A well-established rat model of diabetes was found to harbour mutations in IDE that reduce its ability to degrade both insulin and amyloid- β protein^{6,7}. More recently, genetically modified mice that lack the gene for IDE were created. These mice had elevated insulin levels upon fasting,

as predicted, but they also developed glucose intolerance, and they showed increased levels of cerebral amyloid- β protein⁸. These and other findings suggest that in some cases of diabetes (and perhaps also in some cases of Alzheimer's disease), there might be too little IDE activity rather than too much, with chronically elevated insulin levels perhaps leading to insulin resistance.

If this is true, IDE activators seem to be the logical therapeutic approach, especially for Alzheimer's disease. Current thinking suggests that activators would be difficult to achieve in practice, for the same reason that it is easier to break a machine than to improve its performance. But the work of Shen *et al.*¹ shows that IDE has unorthodox enzymatic properties that might permit activators to be developed after all.

The authors' crystal structures¹ reveal that IDE resembles a clam shell, with two bowl-shaped halves connected by a flexible hinge (Fig. 1). This configuration allows the

protease to switch between 'open' and 'closed' states. Shen *et al.* show that extended hydrogen bonding between the two halves of IDE creates a 'latch' that tends to keep the protease closed (Fig. 1a). Notably, by introducing mutations to the enzyme that destabilize the hydrogen-bond latch, the authors were able to increase the protease's efficiency in cleaving a test substrate by as much as 40-fold (Fig. 1b). This improved efficiency was also seen in the degradation of insulin and amyloid- β protein.

So what is the mechanistic basis of the profound enzyme activation seen in the mutant IDE? This can be understood by considering a simple, two-step model⁹ of the enzyme reaction. First, the enzyme and the substrate bind to each other in a reversible process to form an enzyme–substrate complex. Second, catalytic cleavage of the substrate occurs with concomitant release of the reaction products. Mutations that promote the open state of the protease — thus allowing it to bind substrate — could improve the efficiency of the reaction by accelerating the rate of the enzyme–substrate complex formation.

However, there is a second way that these mutations could activate the protease. In our simple model of the enzyme reaction, the second step actually includes at least two discrete processes: catalysis (that is, substrate cleavage) and dissociation of the products from the enzyme. This complication is usually ignored by assuming that the rate of product dissociation is rapid compared with that of catalysis, making catalysis the rate-limiting step. Although this assumption holds for many proteases, the new work suggests that IDE probably conforms to a more complex kinetic model, where catalysis does not lead automatically to product release. Instead, an additional step is required in which the

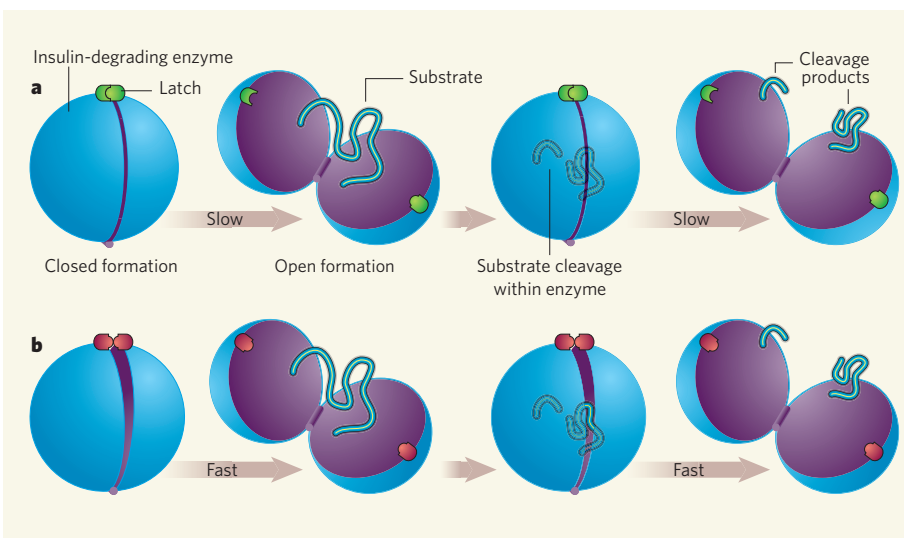


Figure 1 | Enzyme activation. **a**, Insulin-degrading enzyme (IDE) cleaves molecules implicated in diabetes and Alzheimer's disease. The crystal structures of IDE reported by Shen *et al.*¹ reveal a 'latch' mechanism (green) that holds the enzyme in a closed state, delaying entry of the substrate or exit of the cleavage products. **b**, Mutations (red) that disrupt the latch promote the open conformation of the enzyme. Such mutants accept substrates and release products more readily than naturally occurring IDE, and so are more active.

*This article and the paper concerned¹ were published online on 11 October 2006.

protease opens up to allow the products to escape. The findings of Shen *et al.*¹ suggest that this final enzyme-opening step limits the rate of the overall reaction, so that mutations that facilitate this opening process activate IDE.

Shen and colleagues' results¹ excitingly suggest that drug molecules might also be able to activate IDE by mimicking the latch-disrupting mutations. Of course, designing a molecule that targets the latch system may prove challenging. On the other hand, the new structures of IDE will undoubtedly facilitate the development of inhibitors targeting IDE's active site. The potential use of IDE inhibitors for the treatment of diabetes should not be dismissed out of hand, because transient pharmacological inhibition of IDE might yield different results from the life-long, pan-cellular defects present in the genetically altered rodent models discussed above. Moreover, any risk of exacerbating Alzheimer's disease (by elevating amyloid- β -protein levels) might be avoided by using compounds that do not cross into the brain. Whether inhibition or activation of IDE is ultimately prescribed, Shen and colleagues' crystal structures herald the dawn of a rational, structure-based approach to the pharmacology of this ubiquitous protease. ■

Malcolm A. Leissring is in the Department of Biochemistry, Scripps Florida, The Scripps Research Institute, Jupiter, Florida 33458, USA. Dennis J. Selkoe is at the Center for Neurologic Diseases, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mails: leissring@scripps.edu; dselkoe@rics.bwh.harvard.edu

1. Shen, Y., Joachimiak, A., Rosner, M. R. & Tang, W.-J. *Nature* **443**, 823–826 (2006).
2. Simkin, B., Broh-Kahn, R. H. & Mirsky, I. A. *Arch. Biochem.* **20**, 1–9 (1949).
3. Mirsky, I. A. & Perisutti, G. *Science* **122**, 559–560 (1955).
4. Qiu, W. Q. *et al.* *J. Biol. Chem.* **273**, 32730–32738 (1998).
5. Leissring, M. A. *et al.* *Neuron* **40**, 1087–1093 (2003).
6. Fakhrai-Rad, H. *et al.* *Hum. Mol. Genet.* **9**, 2149–2158 (2000).
7. Farris, W. *et al.* *Am. J. Pathol.* **164**, 1425–1434 (2004).
8. Farris, W. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 4162–4167 (2003).
9. Briggs, G. E. & Haldane, J. B. S. *Biochem. J.* **19**, 338–339 (1925).

PALAEOANTHROPOLOGY

Return of the last Neanderthal

Eric Delson and Katerina Harvati

New finds from Gibraltar date Mousterian tools to as recently as 28,000 years ago. By inference, their Neanderthal makers survived in southern Iberia long after all other well-dated occurrences of the species.

The last Neanderthals were participants in one of the most dramatic events in the story of human evolution. At a time of increasing climatic instability and environmental deterioration, they would have had to have survived in ever-smaller groups, confined to less environmentally hostile refugia on the coast of the Mediterranean, and competing for access to resources with modern humans pressing on their territory.

These conditions are widely thought to have led to the Neanderthals' extinction within a relatively short time after the colonization of Europe by modern humans¹. But in a paper² on page 850 of this issue, Finlayson *et al.* revise that model considerably*. They produce dating results from Gorham's Cave, Gibraltar, that might indicate that a group of Neanderthals survived extinction in this part of southern Iberia until at least 28,000 years ago — thousands of years after anatomically modern humans had firmly established themselves as the inheritors of the European continent.

Neanderthals inhabited western Eurasia from a time in the Middle Pleistocene between 500,000 and 160,000 years ago (depending on the definition of the earliest members of the Neanderthal group^{3,4}) until approximately 30,000 years ago (Fig. 1). They were characterized by a suite of specialized morphological features, many of them unique to the group, that together make them highly distinct from modern humans. Their skeletal remains are often found associated with 'Mousterian' stone tools, named after the Le Moustier site

*This article and the paper concerned² were published online on 13 September 2006.

in France. In Europe — but not in northwest Africa or southwest Asia — such tools are exclusively found with Neanderthals, and are presumed to have been made by them.

Neanderthal remains discovered from times near the end of their existence are sometimes found with tool assemblages resembling those produced by early modern humans. This is possibly a result of acculturation or imitation of modern human technology⁵. Although there is still some discussion over the Neanderthals' taxonomic status and their relationship to modern humans, it is now widely recognized that they represent a distinct, Eurasian evolutionary lineage. They shared a common ancestor with modern humans in the early Middle Pleistocene or before^{3,6,7}, but became isolated thereafter from the rest of the Old World. Glacial climatic conditions are considered at least in part responsible for this isolation and for the evolution of some distinctive features of Neanderthal morphology, especially their short limbs and heavy trunks. These are similar to, but more extreme than, features of cold-adapted modern populations such as the Inuit^{8,9}.

The interaction between Neanderthals and modern humans after the arrival of the latter in Europe around 40,000 years ago is among the most interesting topics in European palaeoanthropology. Did they meet? Did they compete? If so, in what ways? Did they interbreed? If they did, did the Neanderthals become assimilated into the modern-human gene pool, or was theirs a union without issue⁴?

Until recently, the interval of coexistence of the two groups in Europe was thought to be as long as 8,000 to 10,000 years. Although it has

ASTRONOMY

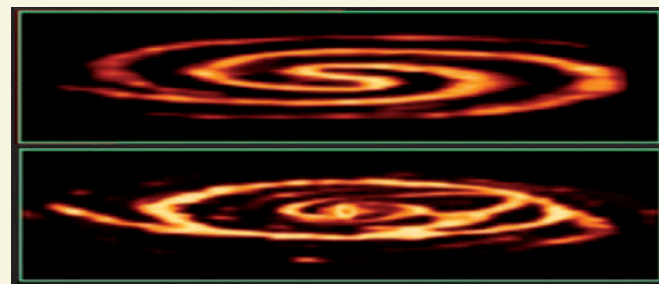
Andromeda's troubled past

It would have been quite a collision. Around 210 million years ago, galaxy M31 — better known as Andromeda — and its smaller neighbour M32 met almost head-on, leaving the battered structure of M31 seen today. This striking proposal is made by D. L. Block *et al.* on page 832 of this issue (*Nature* **443**, 832–834; 2006).

The story is summed up by the two snapshots, each roughly

120,000 light years across, included here. They show the results of Block and colleagues' numerical simulations, with (top) a serene M31 some 35 million years before the collision, and (below) the dishevelled M31 of today.

The simulations offer telling support for the authors' scheme of events, which hinges on two dust rings in the galaxy's disk



(see Fig. 1 on page 832). One, at a radius of about 33,000 light years, is well known; the other, much closer in, was identified only recently with NASA's Spitzer Space Telescope. The two rings emerge from the

simulations — precise in both position and orientation — as a consequence of a head-on galaxy-galaxy impact, with M32 being the likely culprit.

Tim Lincoln

been suggested that modern humans arrived in most areas after Neanderthals had already become locally extinct⁸, most researchers found it difficult to imagine that the two groups never interacted during those millennia¹. The most recent review of the dating evidence¹, however, has proposed that the duration of coexistence was much shorter than thought, and limited to as little as 2,000 years in some places.

Such a short interval points to a relatively fast Neanderthal extinction and a competitive edge for modern humans. That edge came perhaps from cultural practices such as improved clothing and more effective social networks, as documented by finds of personal ornamentation^{9,10}. The short time span also limits the scope for interaction, whether cultural or genetic, between the two groups. This view is consistent both with genetic evidence pointing to very little, if any, Neanderthal contribution to the modern-human gene pool, and with the paucity of potential hybrid skeletal remains.

Finlayson and colleagues' evidence² adds yet another dimension to the story of coexistence. It indicates that at least one group of Neanderthals in Gibraltar was able to survive the deteriorating climate and competition from modern humans, while continuing to use Mousterian (Middle Palaeolithic) technology, until 28,000, and possibly 24,000, years ago. Such a late survival would reinforce the importance of southern Iberia as a refuge area at a time when modern humans were expanding and diversifying culturally across mid-latitude Europe.

Finlayson *et al.* are rightly cautious in their treatment of their very late dates for the last Mousterian occupation at Gibraltar. Their dating sequence (their Fig. 1 on page 851 and Table 1 on page 852) shows several cases where samples lower down in their dig have produced ages that are younger than those above them. This violates the basic rule of stratigraphic superposition, that younger layers overlie older ones. The authors state that such reversals are due to the reuse and cleaning of what is assumed to have been the hearth area, suggesting that dates within the main part of the hearth (which has 24,000 above 26,000 above 30,000 years ago) are more trustworthy than areas alongside it. But there are just too many instances of dates younger than 28,000 years that are out of order, implying that these dates might not be reliable. More extensive sampling of the *in situ* hearth and surroundings might resolve this issue.

If the evidence for Neanderthal survival until 24,000 years ago were to hold water, it might also put other findings in a new light, in particular the Lagar Velho child. This purported Neanderthal-modern hybrid was found in Portugal with Upper Palaeolithic artefacts, and dated¹¹ to 24,500 years ago. Until now, one of the main objections to the acceptance of this find as a possible hybrid — aside from its juvenile morphology, which makes it difficult to draw conclusions about the adult form — has been its chronology. The specimen dates to several millennia after the Neanderthals were thought to



Figure 1 | Principal late Neanderthal/Mousterian sites across western Eurasia. Finlayson and colleagues' find² of Mousterian tools in Gorham's Cave, Gibraltar, might be the most recent indication of Neanderthal settlement yet. Zafarraya (Spain) and Figueira Brava (Portugal) have yielded southern Iberian putative late Neanderthal fossils, whereas St. Césaire and Arcy-sur-Cure (both France) are slightly older (between 35,000 and 30,000 years old). Also indicated are Le Moustier, the eponymous site for the Mousterian tool industry, and Feldhofer, the site of the initial Neander Valley find of 1856. Vindija (Croatia) and Mezmaiskaya (Russia) are noted by Finlayson *et al.*² as having recently been redated to older than 30,000 years ago. Subalyuk (Hungary), Guattari (Italy), Amud (Israel) and Shanidar (Iraq), as well as Feldhofer and Le Moustier, are older than 35,000 years, but they give an indication of the geographical range of the last glacial Neanderthal finds; sites farther east, such as Teshik-Tash (Uzbekistan; not shown), have recently been the subject of questions as to the identity of the fossils recovered.

have disappeared, and was therefore much too recent to be a hybrid¹². This criticism would no longer hold if Finlayson and colleagues' youngest dates could be accepted; for the moment, the consensus is that they cannot.

When considering Finlayson and colleagues' find, it is also important to point out that evidence of Mousterian tools does not in itself indicate that their makers were Neanderthals; this is merely a reasonable assumption. Recovering human fossils from the site would resolve the uncertainty, and possibly shed light on the morphology of late Neanderthals. Until then, we can only be certain that Mousterian toolmakers occupied Gorham's Cave as late as 28,000 years ago. The interval between 28,000 and 24,000 years ago, and the toolmakers' identity, remain in question. Finally, the authors correctly dismiss now discredited late Neanderthal dates from central and eastern Europe, but they may have too quickly rejected some from southern Iberia¹³ that could complement their own evidence for regional refugia.

The late survival of an archaic hominin is not completely unexpected, given the spectacular find of the even later *Homo floresiensis* in Indonesia¹⁴. However, as exciting as the new Gibraltar dates might be, this is not the first claim of the discovery of the last Neanderthal¹⁵. Finlayson and colleagues' date of 28,000 years ago is later than any other well-documented supposed Neanderthal occurrence, proving the importance of this long-term project. Gorham's Cave might well pinpoint the newly extended end of a long lineage of human occupation in Europe. But time will tell.

Eric Delson is in the Department of Anthropology,

Lehman College, Bronx, New York 10468, USA, and at the American Museum of Natural History, New York.

e-mail: eric.delson@lehman.cuny.edu

Katerina Harvati is in the Department of Human Evolution, Max-Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.

e-mail: harvati@eva.mpg.de

1. Mellars, P. *Nature* **439**, 931–935 (2006).
2. Finlayson, C. *et al.* *Nature* **443**, 850–853 (2006).
3. Delson, E. & Baab, K. in *McGraw-Hill Encyclopedia of Science and Technology* Vol. 7, 464–478 (in the press).
4. Harvati, K. & Harrison, T. (eds) *Neanderthals Revisited: New Approaches and Perspectives* (Springer, Dordrecht, in the press).
5. Harrold, F. B. in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans* (eds Mellars, P. & Stringer, C.) 677–713 (Princeton Univ. Press, 1989).
6. Arsuaga, J. L., Martinez, I., Gracia, A. & Lorenzo, C. *J. Hum. Evol.* **33**, 219–281 (1997).
7. Harvati, K., Frost, S. R. & McNulty, K. P. *Proc. Natl Acad. Sci. USA* **101**, 1147–1152 (2004).
8. Finlayson, C. *Neanderthals and Modern Humans* (Cambridge Univ. Press, 2004).
9. Aiello, L. C. & Wheeler, P. in *Neanderthals and Modern Humans in the European Landscape During the Last Glaciation* (eds van Andel, T. H. & Davies, W.) 147–166 (McDonald Inst., Cambridge, UK, 2003).
10. Stringer, C. *et al.* in *Neanderthals and Modern Humans in the European Landscape During the Last Glaciation* (eds van Andel, T. H. & Davies, W.) 233–240 (McDonald Inst., Cambridge, UK, 2003).
11. Duarte, C. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 7604–7609 (1999).
12. Tattersall, I. & Schwartz, J. H. *Proc. Natl Acad. Sci. USA* **96**, 7117–7119 (1999).
13. Hublin, J. J. & Bailey, S. E. in *When Neanderthals and Modern Humans Met* (ed. Conard, N. J.) 105–128 (Kerns, Tübingen, 2006).
14. Brown, P. *et al.* *Nature* **431**, 1055–1061 (2004).
15. Smith, F. H., Trinkaus, E., Pettit, P. B., Karavanic, I. & Paunovic, M. *Proc. Natl Acad. Sci. USA* **96**, 12281–12286 (1999).

BRIEF COMMUNICATIONS

Three-dimensional miniature endoscopy

A single optical fibre acts as a flexible probe to transmit a superior image of an internal landscape.

Endoscopes help medical procedures to be less invasive, thereby reducing the risk of complications as well as costs and recovery times, but their application is limited in part by their size and inflexibility and by their inability to provide a three-dimensional perspective. Here we describe a new type of endoscopy that enables video-rate, three-dimensional images to be transmitted from flexible probes that are comparable in diameter to a human hair. This technology opens up the possibility of moving operations to an outpatient setting, reducing requirements for anaesthesia, and minimizing tissue damage.

The first endoscope was invented almost 50 years ago and consisted of a bundle of optical fibres¹. Miniature endoscopes still use bundles of optical fibres to transmit a two-dimensional image, but larger endoscopes now employ solid-state, charge-coupled-device cameras for superior image quality. Fibre-bundle endoscopes with submillimetre diameters have been used for a variety of clinical applications^{2–5}. However, they have not been widely adopted owing to their rigidity and inadequate image quality, which is due in part to relatively low numbers of pixels and superimposition of a honeycomb pattern (a pixelation artefact).

Spectrally encoded endoscopy (SEE) is a miniature-endoscopy technique that overcomes many of the limitations of fibre-optic imaging bundles⁶. With SEE, polychromatic light emanating from a single optical fibre is configured such that each colour (wavelength) is projected to a different location on the tissue surface⁷ (Fig. 1a). Light reflected from the patient can then be decoded outside the body by using a spectrometer, to form one line of the endoscopic image. The second dimension is obtained by moving the fibre using a mechanical-transduction mechanism, such as an external motor or a galvanometer. The diameter of the SEE probe can be as small as that of the optical fibre, which is typically in the range 80–250 μm . Even though the probe diameter is small, the number of pixels in a SEE image is larger than that from fibre bundles, being dependent only on the spectral width of the light source and the ability of the probe to separate out the components of different wavelengths. Spectral encoding can also provide depth information by using optical interferometry⁸.

Photographs of a prototype miniature SEE probe are shown in the insets of Fig. 1a. Light

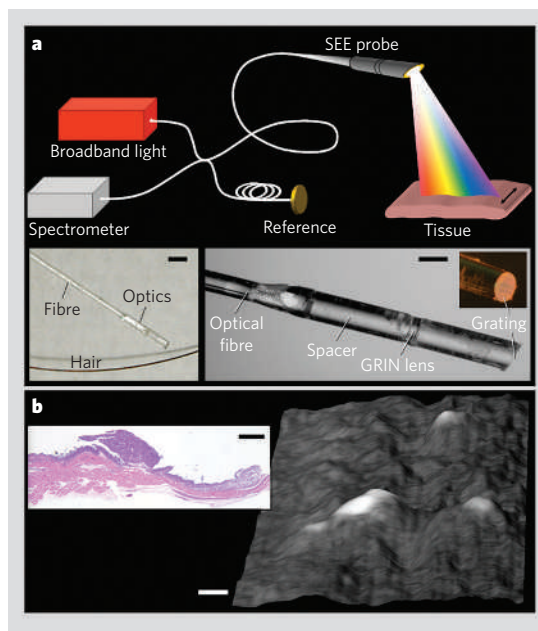


Figure 1 | Miniature, three-dimensional endoscopy through a single optical fibre. **a**, The spectrally encoded endoscopy (SEE) system and probe; tissue is scanned orthogonally to the wavelength axis by rotating the probe (double-headed arrow indicates scan directions). Insets: left, SEE probe next to a human hair, shown for size comparison (scale bar, 1 mm); right, distal optics, and the transmission diffraction grating (scale bar, 0.35 mm). **b**, Image of mouse ovarian tumour nodules on the parietal peritoneal wall *in vivo*. The reflectance image, shown in grey shades (bright greys indicate high reflectance), is superimposed on a three-dimensional rendering of the tissue surface. Inset, cross-sectional histological section (haematoxylin and eosin stain; original magnification, $\times 40$) from the corresponding area. Scale bars, 0.5 mm.

from a single-mode fibre, expanded through a 1.8-mm-long silica spacer, is focused by a gradient-index (GRIN) lens and then diffracted by a transmission grating (1,000 lines per mm) fabricated on the tip of the probe at Littrow's angle. The maximum diameter of the probe is 350 μm . In combination with a broad-bandwidth (700–900 nm) titanium–sapphire laser, Michelson interferometer and high-speed spectrometer (see supplementary information), this miniature probe can be used to obtain volumetric images with about 400,000 resolvable points at video rates (30 frames per second).

To demonstrate the potential of SEE for minimally invasive applications, we imaged metastatic ovarian tumour nodules⁹ on the peritoneum of a living mouse (for methods, see supplementary information). The SEE probe was delivered through a modified 23-gauge needle into the abdominal cavity. A three-dimensional surface image of the parietal peritoneal wall, obtained with the SEE probe *in vivo*, is shown in Fig. 1b: several raised tumour nodules are evident and the inset shows a histological section.

The size and flexibility of this device will allow safer navigation through delicate intraluminal structures such as the fallopian tube, the salivary, mammary and pancreatic ducts, and safer fetal, pediatric and neurosurgical interventions. Navigation mechanisms are still needed that add minimal bulk without

compromising flexibility. Colour could be introduced by using three separate broadbandwidth sources, each centred at visible red, green or blue wavelengths. SEE using fluorescent light should also be useful¹⁰ for imaging labelled molecular species, reporters and molecular beacons.

D. Yelin, I. Rizvi, W. M. White, J. T. Motz, T. Hasan, B. E. Bouma, G. J. Tearney

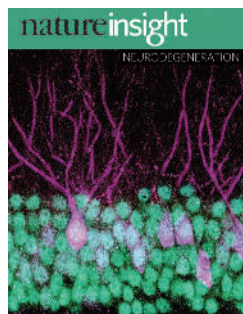
Harvard Medical School and Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
e-mail: gtearney@partners.org

- Hirschowitz, B. I. *Lancet* **1**, 1074–1078 (1961).
- Ishibashi, F., Aziz, K., Abela, G. S. & Waxman, S. J. *Interv. Cardiol.* **19**, 17–25 (2006).
- Jacobs, V. R., Kiechle, M., Plattner, B., Fischer, T. & Paepke, S. J. *Minim. Invas. Gynecol.* **12**, 359–364 (2005).
- Matsunaga, T., Kawakami, Y., Namba, K. & Fujii, M. *Cancer* **101**, 2164–2169 (2004).
- Reece, E. A. *Semin. Perinatol.* **23**, 424–433 (1999).
- Tearney, G. J., Shishkov, M. & Bouma, B. E. *Opt. Lett.* **27**, 412–414 (2002).
- Shi, K. B., Li, P., Yin, S. Z. & Liu, Z. W. *Opt. Expr.* **12**, 2096–2101 (2004).
- Yelin, D., Yun, S. H., Bouma, B. E. & Tearney, G. J. *Opt. Lett.* **30**, 1794–1796 (2005).
- Molpus, K. L. *et al. Int. J. Cancer* **68**, 588–595 (1996).
- Motz, J. T., Yelin, D., Vakoc, B. J., Bouma, B. E. & Tearney, G. J. *Opt. Lett.* **30**, 2760–2762 (2005).

Supplementary information accompanies this communication on Nature's website.

Received 26 July; accepted 14 September 2006.

Competing financial interests: declared none.
doi:10.1038/443765a

**Cover illustration**

Hippocampal granule cells (nuclei, green) lose expression of calcium- and activity-regulated proteins such as calbindin (pink) in a mouse model of Alzheimer's disease. (Courtesy of L. Mucke.)

Editor, *Nature*

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Production Editor

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Jane Walker

Sponsorship

Claire Hines
Claudia Banks

Production

Susan Gray

Marketing

Katy Dunningham

Editorial Assistant

Laura Shaw

NEURODEGENERATION

On 4 November 2006 it will be 100 years since Alois Alzheimer gave his first lecture on a patient with “a peculiar disorder of the cerebral cortex”. The late Auguste D. had suffered from dementia during the last years of her life, and in her brain Alzheimer had discovered protein plaques, neurofibrillary tangles and atherosclerotic changes. These features came to define what is now known as Alzheimer's disease.

Since then, we have learned much about a wide variety of neurodegenerative disorders, and there seem to be some remarkable parallels between them. It is clear that plaques and other atypical protein assemblies are a hallmark of several such disorders, including Parkinson's disease and Huntington's disease, but their significance is still hotly debated. In addition, the failure of neuronal networks and the death of neurons have been implicated in the pathogenesis of various neurodegenerative disorders, as have certain cellular processes such as protein degradation and mitochondrial biology. And, although prion disease may seem to be the odd one out owing to its infectious nature, emerging evidence suggests that several of its peculiarities may also feature in Alzheimer's disease.

We hope that this Insight will offer the reader an exciting, up-to-date guide to the mechanisms and themes that underlie the development of different neurodegenerative diseases. Some of the views expressed may be controversial, but as Alzheimer said, “excessive reservations and paralysing despondency do not help the sciences to advance”. What is needed is “a healthy optimism that cheerfully searches for new ways to understand, as it is convinced that it will be possible to find them”.

We are indebted to our authors, reviewers and advisors for their considerable efforts in producing a stimulating collection of reviews on various aspects of neurodegenerative disease.

Marie-Thérèse Heemels
Senior Editor

REVIEWS

768 **A network dysfunction perspective on neurodegenerative diseases**

J. J. Palop, J. Chin & L. Mucke

774 **A century-old debate on protein aggregation and neurodegeneration enters the clinic**

P. T. Lansbury & H. A. Lashuel

780 **The roles of intracellular protein-degradation pathways in neurodegeneration**

D. C. Rubinsztein

787 **Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases**

M. T. Lin & M. F. Beal

796 **Cell death in the nervous system**

D. E. Bredesen, R. V. Rao & P. Mehlen

803 **Prions and their partners in crime**

B. Caughey & G. S. Baron

nature
insight

A network dysfunction perspective on neurodegenerative diseases

Jorge J. Palop¹, Jeannie Chin¹ & Lennart Mucke¹

Patients with Alzheimer's disease or other neurodegenerative disorders show remarkable fluctuations in neurological functions, even during the same day. These fluctuations cannot be caused by sudden loss or gain of nerve cells. Instead, it is likely that they reflect variations in the activity of neural networks and, perhaps, chronic intoxication by abnormal proteins that the brain is temporarily able to overcome. These ideas have far-reaching therapeutic implications.

The prevalence of neurodegenerative diseases is increasing rapidly¹. Although the speed of research has increased, clinical trials have yielded mostly disappointing results. Are we missing something critical? Are there untested therapeutic entry points that deserve consideration? We propose that reversible network dysfunction may be such an entry point (Figs 1, 2). Although progressive neuronal loss is a hallmark of neurodegenerative disorders (see page 796), some neurological impairment may reflect dysfunction rather than loss of neurons. Abnormal protein assemblies seem to trigger vicious cycles of aberrant neuronal activity and compensatory alterations in neurotransmitter receptors and related signalling pathways that lead to synaptic deficits, disintegration of neural networks, and, ultimately, failure of neurological functions. At least in animal models, some pathogenic cascades can be prevented or reversed by removing abnormal proteins or pharmacologically modulating neuronal activities, interestingly without obvious effects on neuronal number. Enhancing neuronal plasticity might help the remaining neural circuits to compensate for lost or broken circuits and improve overall network performance and neurological function. Improving network activity may also help to prevent the inexorable loss of neuronal processes and cell bodies that occurs in Alzheimer's disease (AD) and other neurodegenerative disorders.

Functional fluctuations and neurological decline

Most clinicians and family members caring for people with neurodegenerative disorders are well aware of the striking fluctuations in functional abilities that these patients experience, often within the same day and covering a surprising range of functional (dis)abilities. The extent of these fluctuations is illustrated by the following caregiver descriptions of patients' best and worst performances within the same day². One carer stated that although at times the patient "was nonsensical, confused, and mumbled incoherently," at other times, "she was almost as she was." And the wife of another patient reported that at one point, "he kept looking for the exit, couldn't find the bedroom or the bathroom and had trouble recognizing me," whereas, at another, "he was alert, opened the door, and greeted me after work. He knew me and seemed pleased to see me." Surprisingly, relatively few studies have systematically investigated this fascinating phenomenon. The few methods available to analyse these fluctuations include a scale for clinicians (Clinician Assessment of Fluctuation) and a scale for non-clinicians (One Day Fluctuation Assessment Scale). Results of these scales correlate significantly with each other as well as with neuropsychological and electrophysiological measures^{3,4}.

Many fluctuations recorded in AD relate to memory, whereas fluctuations in dementia with Lewy bodies (DLB) affect a wider range of functions and tend to be more extreme^{2,4}. Fluctuating cognition with pronounced variations in attention and alertness are so characteristic of DLB that they have been made a core feature in the clinical diagnosis of this condition⁵. Why fluctuations are more severe in DLB than AD is not known. It is also not clear whether and to what extent these fluctuations overlap mechanistically with delirium, which occurs at increased frequency in people with dementias, even in the absence of other medical problems⁶.

Underlying mechanisms

Because the progressive neurological decline seen in neurodegenerative disorders is associated with neuronal loss, the peak performance of patients at any disease stage may be capped by the loss of neurons. However, the extent to which neurological deficits in such conditions relate directly to neuronal loss is controversial^{7–11}. Moreover, similar functional impairments can be caused by processes that do not involve significant neuronal loss^{12,13}.

It is unlikely that changes in neuronal number account for rapid and reversible fluctuations in neurological functions. Such fluctuations probably reflect complex adjustments in molecules, signalling cascades, synaptic modifications, neuronal activities and network interactions.

Transgenic mouse models expressing abnormal proteins associated with AD, DLB, Parkinson's disease (PD), Huntington's disease (HD) and tauopathies such as frontotemporal dementia (FTD) develop distinct disease-related neurological impairments (Box 1). In transgenic mice,

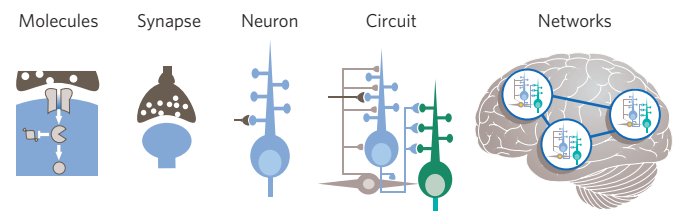


Figure 1 | Neurodegenerative disorders affect neural activities at many levels. Neurodegenerative disorders can disrupt molecular pathways, synapses, neuronal subpopulations and local circuits in specific brain regions, as well as higher-order neural networks. Abnormal network activities may result in a vicious cycle, further impairing the integrity and functions of neurons and synapses, for example, through aberrant excitation or inhibition.

¹Gladstone Institute of Neurological Disease and Department of Neurology, University of California, San Francisco, California 94158, USA.

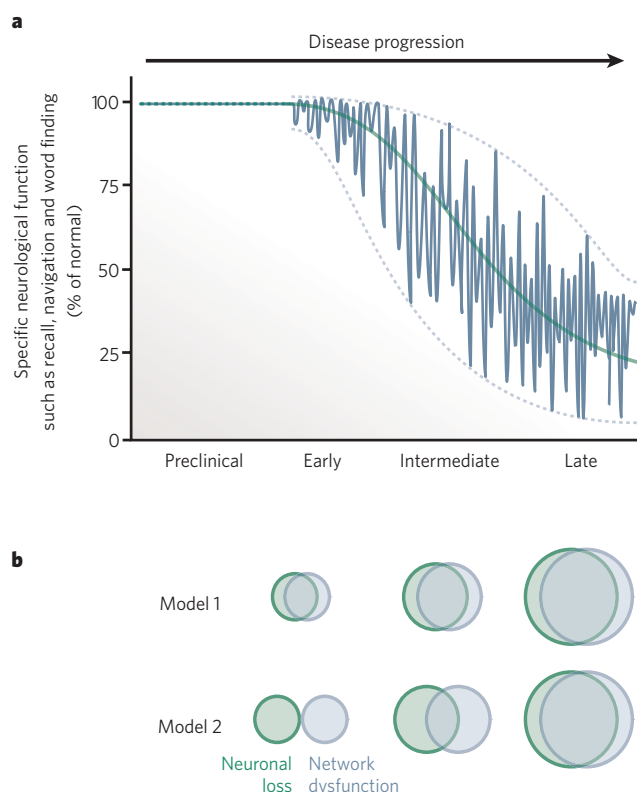


Figure 2 | Functional fluctuations in neurodegenerative disorders may represent fruitful entry points for mechanistic investigations and clinical trials. **a**, Hypothetical diagram contrasting the relatively rapid functional fluctuations (grey) with the more gradual and prolonged overall neurological decline (green). Understanding the mechanisms that separate the peaks from the troughs of fluctuation might allow the development of drugs to optimize performance and minimize disability at different stages of these conditions. The efficacy of such drugs could be assessed much more rapidly than that of drugs aimed at the gradual decline. If there is mechanistic overlap between the fluctuations and the gradual decline, drugs aimed at the fluctuations may also stall disease progression. **b**, The interdependence between network dysfunction and neuronal loss in neurodegenerative disorders remains to be determined. It could be extensive from the start (model 1) or increase as the diseases progress (model 2).

the elimination of abnormal proteins can reverse neurological deficits without changing neuronal number (Box 1). Thus, some neurological impairments that are associated with neurodegenerative conditions might be caused by neuronal dysfunction rather than neuronal loss. This concept is less radical than it might seem at first glance, given that effective neural plasticity allows the brain to cope surprisingly well with even striking neuronal losses^{14,15}.

Many factors might influence what different people can accomplish with their surviving neurons at different stages of their illness. These include the functional state of neurons in affected areas¹⁶, the extent to which alternative neural networks can compensate for lost ones^{17,18}, whether people are able to use specific learning strategies to overcome deficits^{19,20}, genetic factors such as apolipoprotein-E isoforms²¹, and comorbidities such as vascular disease²². Several, if not all, of these modulators represent potential targets for therapeutic intervention.

Even in the absence of disease, neural systems are characterized by extensive 'degeneracy' — in other words, the ability of structurally different elements to perform the same function or yield the same output^{23,24}. The degeneracy of nervous-system design may help to explain why so many neurons in the substantia nigra can die before they are missed at the clinical level, and why most neurodegenerative diseases progress so gradually. The complexity of compensatory mechanisms is well illustrated by the multi-level adjustments that occur in the cortical–basal-ganglia–thalamocortical network at different stages of PD

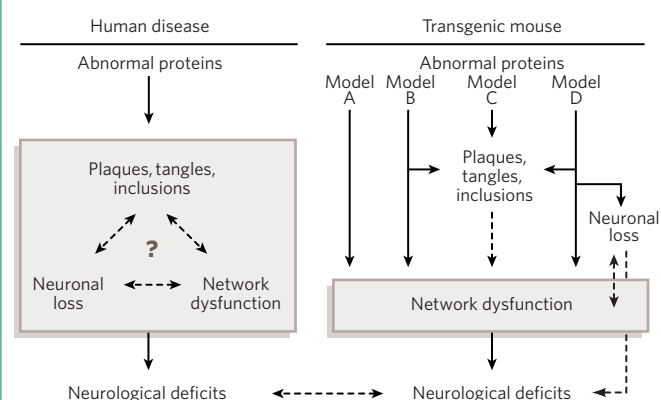
and in related animal models^{16,25}. Progressive failure of these compensatory mechanisms might contribute to the prominent daily fluctuations in the effectiveness (known as the on–off phenomenon) of the drug L-DOPA (L-3,4-dihydroxyphenylalanine) that are often seen in advanced PD.

Thus, within the boundaries set by neuronal loss, the remainder of the nervous system has a notable, albeit variable, capacity to maintain neurological functions. Tremendous benefits might be reaped from therapeutic interventions that maximize a patient's opportunity for optimal performance — shifting the focus from neurons that are lost to those that survive.

From synaptic dysfunction to network failure

Neural plasticity is highly adaptive during both health and disease. It allows animals to interact effectively with their environment and to cope better with neural injuries. A major component of neural plasticity is

Box 1 | Insights from neurodegenerative disease models



In neurodegenerative disorders, the accumulation of abnormal proteins (see page 774), network dysfunction and neurological deficits are associated with pathological hallmarks (plaques, tangles and inclusion bodies) and neuronal loss. Transgenic mouse models can be used to test specific hypotheses about the relationships among these features (see panel, above). The studies highlighted in the table below most strongly support models B and D, demonstrating that some neuronal dysfunction and neurological deficits caused by the abnormal protein assemblies can be independent of plaques, tangles, inclusions and neuronal loss. They also illustrate that reducing the production, enhancing the removal, or neutralizing the actions of abnormal protein assemblies effectively reverses neurological deficits.

Diseases/models	Associated neurological deficits	Prevented or reversed by
Alzheimer's disease		
APP transgenic mice ⁷²	Learning/memory	Active immunization with A β peptides ^{73,74}
		Intraperitoneal injection of anti-A β antibody ^{75,76}
Huntington's disease		
Huntingtin transgenic mice ^{36,37}	Locomotor	Downregulation of transgene expression ^{77,78}
Lewy body diseases including DLB and PD		
α -Synuclein transgenic mice ³⁶	Locomotor	Active immunization with α -synuclein peptide ⁷⁹
Parkin-knockout mice ³⁶	Locomotor	None known
Tauopathies		
Tau transgenic mice ⁸⁰	Learning/memory	Downregulation of transgene expression ⁸¹

APP, amyloid precursor protein.

Box 2 | Mechanisms of neuronal dysfunction in AD models

Normally, neurons are hardy, long-lived cells that adapt swiftly to molecular alterations in their immediate environment and to changes in the activity of the networks they connect with. A common feature of neurodegenerative diseases is the insidious tendency of abnormal protein assemblies to disable mechanisms underlying this remarkable plasticity. Experimental models have helped to identify several molecules that may be involved in this pathogenic process. Although these examples focus on AD and A β , mechanisms of neural plasticity are also disrupted in HD and related conditions.

The list below provides details on the structures, molecules and processes that have been implicated in the pathogenesis of AD.

Activity-regulated gene expression

- Immediate-early genes including those encoding Fos and activity-regulated cytoskeletal protein^{45,47,82–84}

Cell-surface receptors

- Integrins and cell-adhesion molecules^{32,40,85}
- Metabotropic glutamate receptors⁸⁶
- Nicotinic acetylcholine receptors^{40,82,87,88}
- N-methyl-D-aspartate (NMDA) receptors^{47,67,87}

Neurotransmitter release

- Synaptic vesicle cycling or exocytosis⁸⁹

Signalling cascades

- Kinase pathways including the mitogen-activated protein kinase superfamily, cyclin-dependent kinase 5 and tyrosine kinases^{40,67,82,90–92}
- Phosphatases such as calcineurin and striatal-enriched phosphatase^{67,82,87}
- Scaffolding molecules that regulate signalling at the postsynaptic density^{93,94}

Synaptic integrity

- Presynaptic terminals^{95,96}
- Postsynaptic dendritic spines⁹⁷

Synaptic transmission and plasticity

- Basal synaptic transmission^{95,98}
- Long-term potentiation^{26,99,100}
- Paired-pulse modification²⁶

found in synapses, which are actively strengthened and weakened by complex processes to form the dynamic neural circuits and higher-order networks that store memories, give rise to thoughts and harbour the very essence of who we are. Notably, the abnormal proteins that are suspected of causing neurodegenerative disorders impair the integrity or function of presynaptic terminals and postsynaptic specializations^{21,26,27} (Boxes 1, 2). Many mechanisms may be involved, including excitotoxicity^{28,29}, inflammation³⁰, oxidative stress³¹ and other processes^{27,32–34}.

In AD, synaptic loss exceeds neuronal loss, and depletion of synapses and synaptic proteins correlates better with cognitive decline than does the abundance of plaques or tangles; comparable findings have also been made in other neurodegenerative conditions^{16,35–37}.

Chronic alterations in synaptic plasticity and neurotransmission can affect activity-dependent signalling and gene expression, resulting in the disintegration of neural networks and, ultimately, the failure of neural functions (Box 2; Figs 1, 3). Conversely, environmental stimulation, which increases synaptic plasticity, can both delay and decrease pathological alterations, at least in mouse models^{37,38}.

Although neural plasticity may account, at least in part, for the fact that neurodegenerative disorders typically do not become apparent until old age, these disorders might affect plasticity itself³⁹. By eroding the nervous system's main coping mechanism, this 'degeneration of degeneracy' may be a particularly insidious aspect of these conditions, reminiscent of viruses that paralyse the immune mechanisms that the host would use to fend them off.

The lipid carrier apolipoprotein E (apoE) modulates neural plasticity and susceptibility to the most common neurodegenerative disorders. Diverse neural injuries elicit effective plasticity responses with the protective apoE2 and apoE3 isoforms, but not with apoE4, which increases the risk and accelerates the onset of AD, PD and other neurological conditions²¹.

Mechanisms that promote or disrupt neural plasticity might be attractive therapeutic targets. However, predicting the potential value of such targets becomes increasingly difficult as one moves away from the most proximal genetic or environmental triggers and down the branching molecular cascades that connect these triggers to neurological decline. A better understanding of the complexities encountered along this path might facilitate the development of more effective treatments.

Complexities in space and time

The same molecular process or drug activity can have very different functional consequences in different cell types. Although this might seem trivial, it is easy to underestimate its importance. For example, amyloid- β (A β) binds various cell-surface molecules, including neurotransmitter receptors⁴⁰. The network consequences of this depend on whether A β activates or blocks the receptors, and on the cell type on which the receptors are expressed in relevant regions of the AD brain. Overstimulation of a neurotransmitter receptor on excitatory principal cells might result in excitotoxicity, whereas stimulation of a similar receptor on inhibitory interneurons could shut the network down. Overstimulation of such receptors on both types of cell might ignite an ongoing battle between excitatory and inhibitory activities and destabilize the network.

The situation is made still more complex by the expression and modulation of receptors on synaptic versus extrasynaptic sites on the same neuron. The fact that ageing and neurodegenerative disorders can both alter the distribution of neurotransmitter receptors and other molecules makes it even harder to predict the outcome of drug treatments in these diseases. The clinical relevance of these considerations is highlighted by current questions about the wisdom of blocking N-methyl-D-aspartate (NMDA) receptors in AD, which is the main effect of the latest drug approved for the treatment of this illness⁴¹.

Neural networks contain a variety of glial cells with which neurons have complex reciprocal interactions. Notably, abnormal proteins can impair neurons indirectly through pathogenic glial loops involving the production of neurotoxic factors by microglia^{42,43} or the impairment of astroglial support functions⁴⁴.

Complexities can also arise at the level of network interactions (Fig. 3a, b). The radiological, electrophysiological and molecular mapping of neuronal activities across brain regions is beginning to define the functional topology of neurodegenerative disorders, which extends well beyond the areas showing typical pathological alterations^{18,45,46}.

Discrepancies between neuropathological and functional alterations are perhaps not surprising, because some structural alterations may have little functional consequence, some functional alterations may have no detectable structural correlates, and alterations in one brain region can indirectly affect neuronal functions in other regions through efferent or reciprocal connections. For example, although granule cells in the dentate gyrus typically resist neurodegeneration in AD, they can show considerable A β -dependent molecular and functional alterations, possibly reflecting changes in the entorhinal cortex and basal forebrain, from which granule cells receive afferent inputs^{45,47,48}. Thus, the differential vulnerability of neuronal populations to functional impairments is probably determined by both cell-autonomous effects of pathogenic factors and alterations in the activity of the networks in which the vulnerable neurons participate.

Pathogenic processes almost always trigger compensatory mechanisms (Fig. 3). The recognition of an abnormality as a compensatory change, as opposed to a (co)pathogenic one, is crucial, because treatments aimed at its reversal might worsen rather than ameliorate the disease. A particularly pertinent example is the recent demonstration

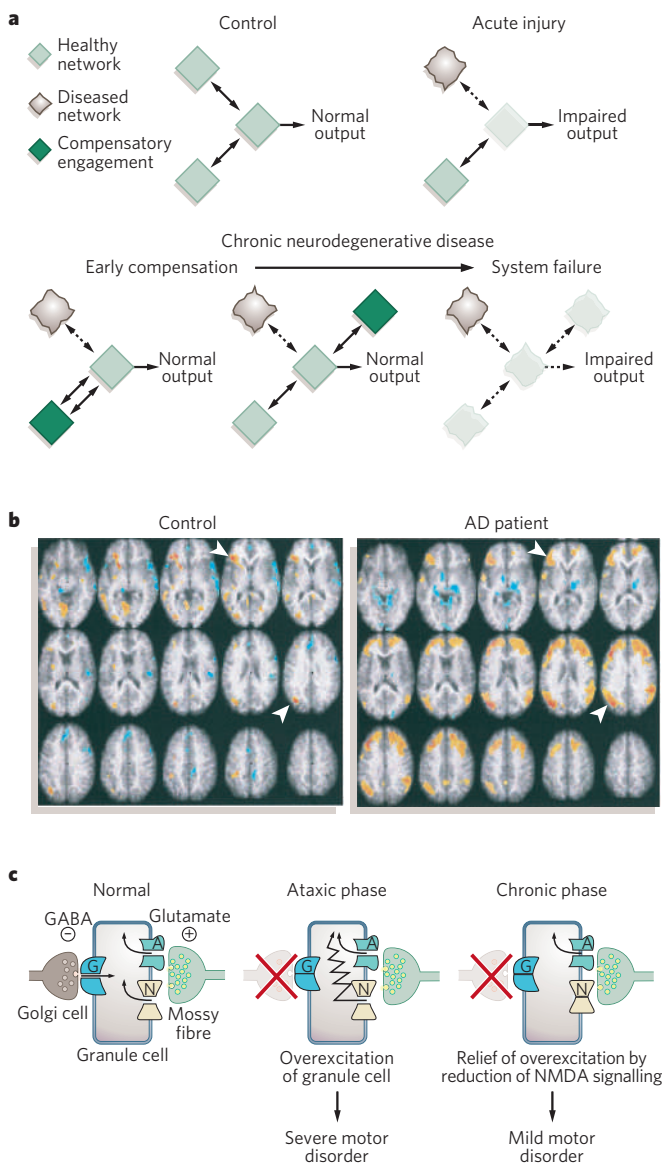


Figure 3 | Neural plasticity counteracts neurodegenerative disorders at multiple levels. **a**, In the healthy brain, activities in different regions (squares) are balanced for optimal overall network performance. Acute injuries to a specific brain region can destabilize the network, upset the activity in other brain regions, and result in neurological deficits. The chronicity of neurodegenerative diseases allows the brain to engage compensatory mechanisms — for example, increasing neuronal activity in undamaged brain regions that normally participate in affected functions and drawing brain regions into the network that are usually involved in other functions. Over time, however, these compensatory mechanisms can fail, and some may even become co-pathogenic, for example, through the excessive stimulation or inhibition of vulnerable neuronal populations. **b**, Positron emission tomography images from a control subject (left) and a patient with early AD (right), illustrating the activation of additional brain regions in the AD case in response to a semantic and an episodic memory task⁶⁹. Whereas controls demonstrated fairly lateralized left-side activity in prefrontal and temporal cortices, AD patients engaged a network involving the bilateral dorsolateral prefrontal and posterior cortices (arrows). Importantly, *de novo* engagement of these networks in AD patients correlated with better performance on memory tasks (pseudocolour indicates brain areas with significant positive correlation with high performance), suggesting that such recruitment of additional areas represents a compensatory mechanism. Comparable activation patterns have also been observed with functional magnetic resonance imaging⁷⁰. Equally pertinent to our network perspective is the demonstration that cognitive decline in AD is associated with a progressive inability to suppress the activation of task-irrelevant brain regions during visual navigation⁴⁶. (Image reproduced, with permission, from ref. 69.) **c**, A particularly instructive example of injury-responsive neural adaptation has been described by Watanabe and colleagues⁵¹. γ -Aminobutyric acid (GABA)-containing inhibitory interneurons in the cerebellum of mice were ablated through an immunotoxin approach. Before and after this manipulation, the investigators determined the time it took for mice to fall off a rotating rod as a measure of coordination. An acute phase of ataxia (incoordination) was followed by a chronic recovery phase during which the lesioned mice were able to stay on the rod if it rotated slowly but not if it rotated rapidly. The images provide an interpretation of results obtained in a detailed analysis of synaptic functions during the different stages of neurological impairment. Elimination of inhibitory interneurons resulted in network dysfunction and ataxia through the unopposed overstimulation of NMDA receptors by glutamate. Network activity was then improved, but not fully normalized, by the compensatory reduction of functional NMDA receptors and the adaptation of synaptic transmission, preventing further overexcitation of cerebellar granule cells. (Adapted, with permission, from ref. 71.) A, α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptor; G, GABA type A receptor; N, NMDA receptor.

that inclusion-body formation in HD does not harm but rather protects neurons against the mutant huntingtin protein⁴⁹.

Some compensatory mechanisms may result in cell survival at the cost of network failure⁴⁵. Vulnerability of excitatory neurons to A β might depend on the amount of inhibitory activity present in a given circuit (known as the circuit buffer)^{47,50}. The compromising nature of injury-responsive synaptic adaptations has been shown elegantly in the neuronal network of the cerebellar cortex⁵¹ (Fig. 3c). This study demonstrates that neural plasticity can result in a significant, albeit partial, recovery of neurological function after destruction of an important neuronal population.

Last, but not least, neurodegenerative diseases are chronic, but not static. For example, choline-acetyltransferase activity in the hippocampus is increased in mild cognitive impairment (MCI), which may precede or represent early AD, but is decreased in later stages of the disease⁵², raising questions about the best timing for cholinergic-replacement therapy in AD⁵³. Radiological imaging studies in patients with MCI have identified hyperactivation of hippocampal and neocortical regions that become hypoactive and atrophic with the progression to AD⁵⁴. Similarly, neuronal activity in the globus pallidus seems to be increased during early stages and decreased in advanced stages of PD²⁵.

To facilitate the development of better treatments for neurodegenerative disorders, these intricacies will have to be confronted with diligence and courage.

Therapeutic implications

From a therapeutic perspective, it is crucial to determine whether the early increases in neural activity are compensatory or pathogenic mechanisms, and to differentiate progression of the original pathogenic process from the emergence of co-pathogens and the age-related failure of compensatory mechanisms (Fig. 3a). If the original trigger of disease must persist for the disease to progress, treatments aimed at the trigger might be effective throughout the course of the disease. If, however, the trigger acts in a 'hit-and-run' event⁵⁵, setting in motion a self-perpetuating cascade, different stages of disease may require distinct therapeutic interventions.

Although early prevention of neuronal loss is clearly an important objective, it is also important to recognize that a proportion of the deficits associated with neurodegenerative disorders might reflect reversible network dysfunction. The therapeutic implications of such dysfunction can be illustrated with reference to epilepsy. Most cases of epilepsy are of unknown and probably multifactorial origin⁵⁶. Although available antiepileptic drugs are not typically categorized as 'disease modifying', they can make an enormous difference in the lives of epileptic patients. Uncontrolled seizures are associated with neurological decline, neuronal loss and premature mortality^{57,58}, and antiepileptic drugs reasonably control seizures in many patients⁵⁶.

Might improvements in network activities in neurodegenerative disorders have similar benefits? At the very least, optimizing the plasticity

of remaining neural networks could allow patients to function as well as possible within the limits of lost neural structures. In the best case, normalization of network activities could help to prevent progressive loss of neurons, because both excessive and inadequate neuronal stimulation can contribute to neurodegeneration.

What will it take to normalize neural plasticity and network activity in neurodegenerative disorders? Although some treatments seem to improve the function of remaining neural circuits, their effect is limited and none seems capable of a permanent functional rescue^{34,59,60}. This disappointing result is somewhat surprising, as the treatments were based on sound rationales relating to the degeneration of particular neuronal populations and well-established deficiencies or imbalances in specific neurotransmitter systems. Perhaps the strategies target too few of the most critical derangements, or the derangements are too advanced or dynamic for standard regimens⁶¹. Multidisciplinary studies are needed to better define the dysfunction in key networks in different neurodegenerative diseases and related animal models.

Interestingly, several neurodegenerative disorders, including AD, DLB and HD, are associated with an increased incidence of seizures^{62–64}. Transgenic mice whose neurons express human amyloid precursor proteins or mutant huntingtin are also prone to epileptic activity^{65,66}, and A β toxicity in some cell-culture models can be inhibited with anticonvulsants²⁸. These and other observations^{29,34} suggest that the relationship between neurodegenerative disorders and epilepsy should be further explored, particularly during early stages of disease and in susceptible subpopulations.

Drugs aimed at neurotransmitter receptors might fail because related signalling pathways are disrupted by abnormal protein assemblies, creating a downstream block that cannot be overcome by upstream modulation. Encouragingly, experimental models are helping to unravel the molecular cascades through which abnormal protein assemblies might cause this disruption (Box 2). Evidence from AD and HD models has implicated abnormal proteins in the disruption of glutamatergic neurotransmission and calcium signalling^{29,33,34,45,47,67}. Further work is needed to determine whether the therapeutic modulation of affected pathways (Box 2) can improve network activities and neurological functions in neurodegenerative conditions and enhance the efficacy of available treatments.

At present, diverse efforts are being directed at the elimination of the abnormal protein assemblies themselves, and with good reason^{27,68}. Reducing the production, enhancing the removal or neutralizing the actions of these assemblies may improve neuronal survival. If the functional decline in neurodegenerative diseases is caused primarily by the gradual loss of neurons, it may take years to detect benefits of these disease-modifying interventions — a daunting prospect that may discourage the development of better drugs for AD and related conditions. However, if the pathogenic proteins actively interfere with synaptic functions and related signalling cascades, lowering their levels or inhibiting their actions might have more immediate effects that could become apparent within weeks or even days (Fig. 2). This hypothesis is supported by results obtained in various experimental models (Box 1).

If future studies confirm the pathogenic importance of reversible network dysfunction in neurodegenerative disorders, it may become possible to shorten clinical-trial periods and to evaluate a greater diversity of therapeutic compounds. This could accelerate the pace of drug validation and offer obvious cost savings. It might also facilitate the identification of effective combinations of drugs with distinct modes of action. Although such regimens are difficult to assess in clinical trials, they have proved useful in other multifactorial diseases (such as epilepsy, hypertension and cancer) and will probably also be required for the effective treatment of neurodegenerative disorders. ■

1. Cowan, W. M. & Kandel, E. R. Prospects for neurology and psychiatry. *J. Am. Med. Assoc.* **285**, 594–600 (2001).
2. Bradshaw, J., Saling, M., Hopwood, M., Anderson, V. & Brodthmann, A. Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. *J. Neurol. Neurosurg. Psychiatry* **75**, 382–387 (2004).
3. Walker, M. P. et al. The clinician assessment of fluctuation and the one day fluctuation assessment scale. Two methods to assess fluctuating confusion in dementia. *Br. J. Psychiatry* **177**, 252–256 (2000).

4. Walker, M. P. et al. Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease, and vascular dementia. *Neurology* **54**, 1616–1625 (2000).
5. McKeith, I. G. et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* **65**, 1863–1872 (2005).
6. Robertsson, B., Blennow, K., Gottfries, C. G. & Wallin, A. Delirium in dementia. *Int. J. Geriatr. Psychiatry* **13**, 49–56 (1998).
7. Terry, R. D. et al. Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* **30**, 572–580 (1991).
8. Morrison, J. H. & Hof, P. R. Life and death of neurons in the aging brain. *Science* **278**, 412–419 (1997).
9. Kordower, J. H. et al. Loss and atrophy of layer II entorhinal cortex neurons in elderly people with mild cognitive impairment. *Ann. Neurol.* **49**, 202–213 (2001).
10. Price, J. L. et al. Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Arch. Neurol.* **58**, 1395–1402 (2001).
11. Greffard, S. et al. Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. *Arch. Neurol.* **63**, 584–588 (2006).
12. Blanchet, P. J. Antipsychotic drug-induced movement disorders. *Can. J. Neurol. Sci.* **30** (Suppl. 1), S101–S107 (2003).
13. Martino, D. & Giovannoni, G. Anticholinergic drugs and their relevance to movement disorders. *Curr. Opin. Neurol.* **17**, 425–432 (2004).
14. Lewin, R. Is your brain really necessary? *Science* **210**, 1232–1234 (1980).
15. Chen, R., Cohen, L. G. & Hallett, M. Nervous system reorganization following injury. *Neuroscience* **111**, 761–773 (2002).
16. Bezard, E., Gross, C. E. & Brotchie, J. M. Presymptomatic compensation in Parkinson's disease is not dopamine-mediated. *Trends Neurosci.* **26**, 215–221 (2003).
17. Stern, Y. What is cognitive reserve? Theory and research application of the reserve concept. *J. Int. Neuropsychol. Soc.* **8**, 448–460 (2002).
18. Buckner, R. L. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron* **44**, 195–208 (2004).
19. Iaria, G., Petrides, M., Dagher, A., Pike, B. & Bohbot, V. D. Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J. Neurosci.* **23**, 5945–5952 (2003).
20. Maguire, E. A., Valentine, E. R., Wilding, J. M. & Kapur, N. Routes to remembering: the brains behind superior memory. *Nature Neurosci.* **6**, 90–95 (2003).
21. Mahley, R. W., Weisgraber, K. H. & Huang, Y. Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **103**, 5644–5651 (2006).
22. Zlokovic, B. V. Neurovascular mechanisms of Alzheimer's neurodegeneration. *Trends Neurosci.* **28**, 202–208 (2005).
23. Prinz, A. A., Bucher, D. & Marder, E. Similar network activity from disparate circuit parameters. *Nature Neurosci.* **7**, 1345–1352 (2004).
24. Edelman, G. M. & Gally, J. A. Degeneracy and complexity in biological systems. *Proc. Natl Acad. Sci. USA* **98**, 13763–13768 (2001).
25. Whone, A. L., Moore, R. Y., Piccini, P. P. & Brooks, D. J. Plasticity of the nigrostriatal pathway in Parkinson's disease. *Ann. Neurol.* **53**, 206–213 (2003).
26. Walsh, D. M. & Selkoe, D. J. Deciphering the molecular basis of memory failure in Alzheimer's disease. *Neuron* **44**, 181–193 (2004).
27. Muchowski, P. J. & Wacker, J. L. Modulation of neurodegeneration by molecular chaperones. *Nature Rev. Neurosci.* **6**, 11–22 (2005).
28. Mark, R. J., Ashford, J. W., Goodman, Y. & Mattson, M. P. Anticonvulsants attenuate amyloid β -peptide neurotoxicity, Ca^{2+} deregulation, and cytoskeletal pathology. *Neurobiol. Aging* **16**, 187–198 (1995).
29. Hynd, M. R., Scott, H. L. & Dodd, P. R. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochem. Int.* **45**, 583–595 (2004).
30. Wyss-Coray, T. & Mucke, L. Inflammation in neurodegenerative disease — a double-edged sword. *Neuron* **35**, 419–432 (2002).
31. Beal, M. F. Mitochondria take center stage in aging and neurodegeneration. *Ann. Neurol.* **58**, 495–505 (2005).
32. Small, D. H., Mok, S. S. & Bornstein, J. C. Alzheimer's disease and A β toxicity: from top to bottom. *Nature Rev. Neurosci.* **2**, 595–598 (2001).
33. Mattson, M. P. Pathways towards and away from Alzheimer's disease. *Nature* **430**, 631–639 (2004).
34. Handley, O. J., Naji, J. J., Dunnett, S. B. & Rosser, A. E. Pharmaceutical, cellular and genetic therapies for Huntington's disease. *Clin. Sci. (Lond.)* **110**, 73–88 (2006).
35. Honer, W. G. Pathology of presynaptic proteins in Alzheimer's disease: more than simple loss of terminals. *Neurobiol. Aging* **24**, 1047–1062 (2003).
36. Levine, M. S., Cepeda, C., Hickey, M. A., Fleming, S. M. & Chesselet, M. F. Genetic mouse models of Huntington's and Parkinson's diseases: illuminating but imperfect. *Trends Neurosci.* **27**, 691–697 (2004).
37. van Dellen, A., Grote, H. E. & Hannan, A. J. Gene-environment interactions, neuronal dysfunction and pathological plasticity in Huntington's disease. *Clin. Exp. Pharmacol. Physiol.* **32**, 1007–1019 (2005).
38. Lazarov, O. et al. Environmental enrichment reduces A β levels and amyloid deposition in transgenic mice. *Cell* **120**, 701–713 (2005).
39. Mesulam, M. M. Neuroplasticity failure in Alzheimer's disease: bridging the gap between plaques and tangles. *Neuron* **24**, 521–529 (1999).
40. Verdier, Y., Zarandi, M. & Penke, B. Amyloid β -peptide interactions with neuronal and glial cell plasma membrane: binding sites and implications for Alzheimer's disease. *J. Pept. Sci.* **10**, 229–248 (2004).
41. Schmitt, H. P. Pouring oil into the fire? On the conundrum of the beneficial effects of NMDA receptor antagonists in Alzheimer disease. *Psychopharmacology (Berl.)* **179**, 151–153 (2005).
42. Chen, J. et al. SIRT1 protects against microglia-dependent amyloid- β toxicity through inhibiting NF- κ B signaling. *J. Biol. Chem.* **280**, 40364–40374 (2005).
43. Giorgini, F., Guidetti, P., Nguyen, Q., Bennett, S. C. & Muchowski, P. J. A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. *Nature Genet.* **37**, 526–531 (2005).

44. Eddleston, M. P. & Mucke, L. Molecular profile of reactive astrocytes — implications for their role in neurologic disease. *Neuroscience* **54**, 15–36 (1993).
45. Palop, J. J. *et al.* Neuronal depletion of calcium-dependent proteins in the dentate gyrus is tightly linked to Alzheimer's disease-related cognitive deficits. *Proc. Natl Acad. Sci. USA* **100**, 9572–9577 (2003).
46. Drzezza, A. *et al.* Impaired cross-modal inhibition in Alzheimer disease. *PLoS Med.* **2**, 986–995 (2005).
47. Palop, J. J. *et al.* Vulnerability of dentate granule cells to disruption of Arc expression in human amyloid precursor protein transgenic mice. *J. Neurosci.* **25**, 9686–9693 (2005).
48. Colom, L. V. Septal networks: relevance to theta rhythm, epilepsy and Alzheimer's disease. *J. Neurochem.* **96**, 609–623 (2006).
49. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805–810 (2004).
50. Graf, R. A. & Kater, S. B. Inhibitory neuronal activity can compensate for adverse effects of β -amyloid in hippocampal neurons. *Brain Res.* **786**, 115–121 (1998).
51. Watanabe, D. *et al.* Ablation of cerebellar Golgi cells disrupts synaptic integration involving GABA inhibition and NMDA receptor activation in motor coordination. *Cell* **95**, 17–27 (1998).
52. DeKosky, S. T. *et al.* Upregulation of choline acetyltransferase activity in hippocampus and frontal cortex of elderly subjects with mild cognitive impairment. *Ann. Neurol.* **51**, 145–155 (2002).
53. Small, D. H. Do acetylcholinesterase inhibitors boost synaptic scaling in Alzheimer's disease? *Trends Neurosci.* **27**, 245–249 (2004).
54. Dickerson, B. C. *et al.* Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD. *Neurology* **65**, 404–411 (2005).
55. Oldstone, M. B. Molecular mimicry, microbial infection, and autoimmune disease: evolution of the concept. *Curr. Top. Microbiol. Immunol.* **296**, 1–17 (2005).
56. Duncan, J. S., Sander, J. W., Sisodiya, S. M. & Walker, M. C. Adult epilepsy. *Lancet* **367**, 1087–1100 (2006).
57. Thompson, P. J. & Duncan, J. S. Cognitive decline in severe intractable epilepsy. *Epilepsia* **46**, 1780–1787 (2005).
58. Forsgren, L. *et al.* Mortality of epilepsy in developed countries: a review. *Epilepsia* **46** (Suppl. 11), 18–27 (2005).
59. Schapira, A. H. Present and future drug treatment for Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **76**, 1472–1478 (2005).
60. Lleo, A., Greenberg, S. M. & Growdon, J. H. Current pharmacotherapy for Alzheimer's disease. *Annu. Rev. Med.* **57**, 513–533 (2006).
61. Gsell, W., Jungkunz, G. & Riederer, P. Functional neurochemistry of Alzheimer's disease. *Curr. Pharm. Des.* **10**, 265–293 (2004).
62. Nance, M. A. & Myers, R. H. Juvenile onset Huntington's disease — clinical and research perspectives. *Ment. Retard. Dev. Disabil. Res. Rev.* **7**, 153–157 (2001).
63. Weiner, M. F. *et al.* Can Alzheimer's disease and dementias with Lewy bodies be distinguished clinically? *J. Geriatr. Psychiatry Neurol.* **16**, 245–250 (2003).
64. Amati, J. C. *et al.* Incidence and predictors of seizures in patients with Alzheimer's disease. *Epilepsia* **47**, 867–872 (2006).
65. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493–506 (1996).
66. Del Vecchio, R. A., Gold, L. H., Novick, S. J., Wong, G. & Hyde, L. A. Increased seizure threshold and severity in young transgenic CRND8 mice. *Neurosci. Lett.* **367**, 164–167 (2004).
67. Xie, C. W. Calcium-regulated signaling pathways: role in amyloid β -induced synaptic dysfunction. *Neuromolecular Med.* **6**, 53–64 (2004).
68. Selkoe, D. J. & Schenk, D. Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **43**, 545–584 (2003).
69. Grady, C. L. *et al.* Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. *J. Neurosci.* **23**, 986–993 (2003).
70. Pariente, J. *et al.* Alzheimer's patients engage an alternative network during a memory task. *Ann. Neurol.* **58**, 870–879 (2005).
71. Nakanishi, S. Synaptic mechanisms of the cerebellar cortical network. *Trends Neurosci.* **28**, 93–100 (2005).
72. Kobayashi, D. T. & Chen, K. S. Behavioral phenotypes of amyloid-based genetically modified mouse models of Alzheimer's Disease. *Genes Brain Behav.* **4**, 173–196 (2005).
73. Janus, C. *et al.* A β peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature* **408**, 979–982 (2000).
74. Morgan, D. *et al.* A β peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature* **408**, 982–985 (2000).
75. Dodart, J. C. *et al.* Immunization reverses memory deficits without reducing brain A β burden in Alzheimer's disease model. *Nature Neurosci.* **5**, 452–457 (2002).
76. Kotilinek, L. A. *et al.* Reversible memory loss in a mouse transgenic model of Alzheimer's disease. *J. Neurosci.* **22**, 6331–6335 (2002).
77. Yamamoto, A., Lucas, J. J. & Hen, R. Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. *Cell* **101**, 57–66 (2000).
78. Martin-Aparicio, E. *et al.* Proteasomal-dependent aggregate reversal and absence of cell death in a conditional mouse model of Huntington's disease. *J. Neurosci.* **21**, 8772–8781 (2001).
79. Masliah, E. *et al.* Effects of α -synuclein immunization in a mouse model of Parkinson's disease. *Neuron* **46**, 857–868 (2005).
80. Brandt, R., Hundelt, M. & Shahani, N. Tau alteration and neuronal degeneration in tauopathies: mechanisms and models. *Biochim. Biophys. Acta* **1739**, 331–354 (2005).
81. Santacruz, K. *et al.* Tau suppression in a neurodegenerative mouse model improves memory function. *Science* **309**, 476–481 (2005).
82. Chin, J. *et al.* Fyn kinase induces synaptic and cognitive impairments in a transgenic mouse model of Alzheimer's disease. *J. Neurosci.* **25**, 9694–9703 (2005).
83. Dickey, C. A. *et al.* Selectively reduced expression of synaptic plasticity-related genes in amyloid precursor protein + presenilin-1 transgenic mice. *J. Neurosci.* **23**, 5219–5226 (2003).
84. Lacor, P. N. *et al.* Synaptic targeting by Alzheimer's-related amyloid β oligomers. *J. Neurosci.* **24**, 10191–10200 (2004).
85. Cotman, C. W., Hailer, N. P., Pfister, K. K., Soltesz, I. & Schachner, M. Cell adhesion molecules in neural plasticity and pathology: similar mechanisms, distinct organizations? *Prog. Neurobiol.* **55**, 659–669 (1998).
86. Wang, Q., Walsh, D. M., Rowan, M. J., Selkoe, D. J. & Anwyl, R. Block of long-term potentiation by naturally secreted and synthetic amyloid β -peptide in hippocampal slices is mediated via activation of the kinases c-Jun N-terminal kinase, cyclin-dependent kinase 5, and p38 mitogen-activated protein kinase as well as metabotropic glutamate receptor type 5. *J. Neurosci.* **24**, 3370–3378 (2004).
87. Snyder, E. M. *et al.* Regulation of NMDA receptor trafficking by amyloid- β . *Nature Neurosci.* **8**, 1051–1058 (2005).
88. Oddo, S. & LaFerla, F. M. The role of nicotinic acetylcholine receptors in Alzheimer's disease. *J. Physiol. (Paris)* **99**, 172–179 (2006).
89. Kelly, B. L., Vassar, R. & Ferreira, A. β -amyloid-induced dynamin 1 depletion in hippocampal neurons. A potential mechanism for early cognitive decline in Alzheimer disease. *J. Biol. Chem.* **280**, 31746–31753 (2005).
90. Haddad, J. J. Mitogen-activated protein kinases and the evolution of Alzheimer's: a revolutionary neurogenetic axis for therapeutic intervention? *Prog. Neurobiol.* **73**, 359–377 (2004).
91. Lee, G. Tau and src family tyrosine kinases. *Biochim. Biophys. Acta* **1739**, 323–330 (2005).
92. Giese, K. P., Ris, L. & Plattner, F. Is there a role of the cyclin-dependent kinase 5 activator p25 in Alzheimer's disease? *Neuroreport* **16**, 1725–1730 (2005).
93. Roselli, F. *et al.* Soluble β -amyloid_{1–40} induces NMDA-dependent degradation of postsynaptic density-95 at glutamatergic synapses. *J. Neurosci.* **25**, 11061–11070 (2005).
94. Almeida, C. G. *et al.* Beta-amyloid accumulation in APP mutant neurons reduces PSD-95 and GluR1 in synapses. *Neurobiol. Dis.* **20**, 187–198 (2005).
95. Hsia, A. *et al.* Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proc. Natl Acad. Sci. USA* **96**, 3228–3233 (1999).
96. Mucke, L. *et al.* High-level neuronal expression of A β _{1–42} in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).
97. Moolman, D. L., Vitolo, O. V., Vonsattel, J. P. & Shelanski, M. L. Dendrite and dendritic spine alterations in Alzheimer models. *J. Neurocytol.* **33**, 377–387 (2004).
98. Kamenetz, F. *et al.* APP processing and synaptic function. *Neuron* **37**, 925–937 (2003).
99. Chapman, P. F. *et al.* Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice. *Nature Neurosci.* **2**, 271–276 (1999).
100. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416**, 535–539 (2002).

Acknowledgments We thank S. Finkbeiner for helpful discussion of the Huntington's disease literature, J. Carroll for preparation of graphics, G. Howard and S. Ordway for editorial review, and D. McPherson and L. Manuntag for administrative assistance. This work was supported by grants (to L.M.) from the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to L.M. (lmucke@gladstone.ucsf.edu).

A century-old debate on protein aggregation and neurodegeneration enters the clinic

Peter T. Lansbury¹ & Hilal A. Lashuel²

The correlation between neurodegenerative disease and protein aggregation in the brain has long been recognized, but a causal relationship has not been unequivocally established, in part because a discrete pathogenic aggregate has not been identified. The complexity of these diseases and the dynamic nature of protein aggregation mean that, despite progress towards understanding aggregation, its relationship to disease is difficult to determine in the laboratory. Nevertheless, drug candidates that inhibit aggregation are now being tested in the clinic. These have the potential to slow the progression of Alzheimer's disease, Parkinson's disease and related disorders and could, if administered presymptotically, drastically reduce the incidence of these diseases. The clinical trials could also settle the century-old debate about causality.

When Alois Alzheimer discovered that the postmortem brains of his severely demented patients contained proteinaceous amyloid plaques, he wondered whether these deposits caused or resulted from neurodegeneration¹. A century later, this question has been vigorously investigated and debated, but not definitively answered. Protein aggregation of some type is now also known to be a characteristic feature of Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), Huntington's disease (HD) and many less prevalent neurodegenerative diseases. However, the fibrillar deposits that are found in postmortem diseased brains — for example, the amyloid plaques noted by Alzheimer, the Lewy bodies of PD and related diseases, and the nuclear inclusions of HD — can be present in the brains of asymptomatic individuals and do not correlate to the severity of disease at time of death. Spurred on by a greater understanding of the *in vitro* protein fibrillization process, a search for a specific neurotoxic prefibrillar aggregate and the pathway by which it causes neuronal dysfunction, death and disease is underway.

The precise relationship between protein aggregates such as the extracellular amyloid plaques and intracellular neurofibrillary tangles that characterize the Alzheimer's disease (AD) brain and the neurodegeneration that produces symptoms has not been determined (Fig. 1). The possibility that protein aggregation causes neurodegeneration (as opposed to resulting from it) returned to centre stage with the discovery that a gene linked to autosomal dominant familial AD encodes a precursor (amyloid precursor protein, APP) of the amyloid- β protein (A β)². A β is a 4-kDa peptide that comprises the least soluble, fibrillar component of AD amyloid plaques^{3,4}. During the past 15 years, a working hypothesis has been developed implicating protein aggregation as the trigger of the cascade of events that result in neurodegeneration and disease⁵. This model, which is known as the 'amyloid hypothesis', has evolved with our understanding of the process of protein aggregation and has withstood concerted efforts to disprove it. In recent years, the hypothesis has been generalized as it has become apparent that most neurodegenerative diseases are characterized by protein aggregates, albeit of varying composition. This review briefly summarizes the key experiments that led to the working hypothesis, but focuses on recent experimental findings that purport to test elements of the working model.

The amyloid hypothesis has been shaped by evidence drawn primarily from three disciplines: neuropathology, genetics and biophysics. Each class of evidence is powerful, but each also has limitations. Pathological studies have focused on defining and refining the correlation between protein deposits, neuronal dysfunction, neuronal loss and severity of symptoms at time of death. Because studies of human tissue rely on postmortem samples, it is difficult to glimpse the early stages of disease during which the triggering species is most likely to be observable. However, this critical period is theoretically observable in living patients by brain imaging. Recent advances towards the *in vivo* detection of

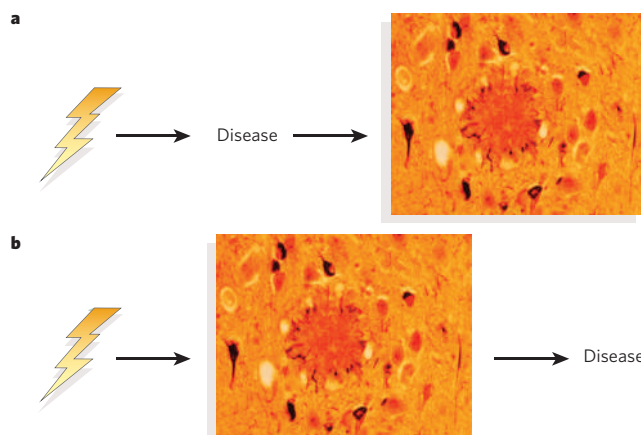


Figure 1 | Fibril formation and disease are linked. Two mutually exclusive hypotheses have been proposed to explain the correlation between neurodegenerative disease and fibrillar-protein deposits in the postmortem diseased brain. **a**, Hypothesis one: human neurodegenerative disease causes fibrillar deposits, but protein aggregation has no causal role. **b**, Hypothesis two: fibrillar protein deposits cause neurodegenerative disease. The tissue image is from an AD brain and shows an amyloid plaque surrounded by intraneuronal neurofibrillary tangles.

¹Department of Neurology, Harvard Medical School and Brigham and Women's Hospital, 65 Landsdowne Street, Cambridge, Massachusetts 02139, USA. Present address: Link Medicine, 790 Memorial Drive, Cambridge, Massachusetts 02139, USA. ²Laboratory of Molecular Neurobiology and Neuroproteomics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

Box 1 | Imaging aggregates may allow presymptomatic diagnosis

It has been estimated that neurodegeneration associated with AD and PD starts 5 to 10 years before symptoms warrant a diagnosis. In the case of PD, imaging studies have shown that loss of 50% of the substantia nigra's dopaminergic neurons is tolerated; symptoms typically appear when 70% have been lost⁸¹. The development of methods to diagnose these diseases presymptomatically has not been a commercial priority, because available therapies target only the symptoms, not the underlying disease. However, this situation is about to change, as the use of therapies that slow the underlying disease (Box 2) will be greatly increased if disease can be diagnosed and treated presymptomatically (Fig. 4). As more genetic risk factors emerge, it will be possible to assess risk, but in order to justify treatment, it will be necessary to directly measure the disease process.

This is possible for PD, because the affected region of the brain is localized and excellent methods for imaging dopaminergic neurons are available. However, in the case of AD, direct imaging of protein aggregates may provide a more sensitive measure. William Klunk of the University of Pittsburgh has developed a positron emission tomography imaging reagent that targets fibrillar amyloid plaques and allows their detection *in vivo*⁶. However, given that fibrillar amyloid occurs relatively late in the disease process, it would also be useful to develop analogous reagents that target early A β -derived aggregates. A recent study in transgenic AD mouse models⁴⁷ used classical biochemical separation methods to correlate a non-fibrillar A β -containing species (A β *56) with the appearance of an AD-like phenotype. An analogous study, correlating an early A β aggregate with the first clinical signs of AD, would require the development of new technology and would have enormous practical consequences. A time can be imagined when individuals with a relative risk of >2.0, on the basis of genetic factors, would be routinely imaged at the age of 50 using such a reagent (Fig. 4). If abnormal A β aggregates were present, treatment with a disease-modifying drug could begin immediately. This approach might effectively 'cure' many cases of AD.

amyloid plaques in human patients with AD^{6–8} mean that it may soon be possible to follow the natural history of protein aggregation and correlate it with disease progression (Box 1). In the meantime, studies of animal models have provided detailed information about certain phenotypes of the human disease^{9,10}.

Genetic data are clearly relevant to human AD and have been instrumental in the building of a circumstantial case to support existing models of pathogenesis. The genes that have been linked to disease are predominantly of two types: familial and susceptibility factors. Several genes of the former type have been identified, and, when mutated, these cause rare, early-onset forms of disease. Examples include the genes encoding APP, which is involved in familial AD, and α -synuclein, which is involved in familial PD. Whether these monogenic subtypes are representative of the vast majority of cases, which are believed to originate from a complex mixture of genetic and perhaps environmental factors, is debatable (see below). The latter type of disease–gene link is based on a susceptibility factor, such as apolipoprotein-E genotype in AD¹¹ or UCH-L1 genotype in PD¹². Although these susceptibility factors are found in many patients, it is difficult to translate their subtle effects into a working hypothesis.

Finally, biophysical studies have been increasingly influential in this field owing to their ability to provide a mechanistic rationale to better explain the effects of disease-causing mutations — for example, why the A β 42/A β 40 ratio is more critical than the total amount of A β produced¹³. The obvious shortcoming of such studies is that the idealized experimental systems that they require bear little resemblance to the human brain. Moreover, the pathological processes are difficult to study *in vitro*, because, unlike biologically necessary processes (such as the complement cascade) or structures (for example, the proteasome), protein aggregation has not been optimized by evolution and is difficult to reconstitute in an artificial environment (in fact, evolution has probably done its best to reduce protein aggregation).

This review discusses and evaluates a series of increasingly detailed hypotheses that seek to explain the relationship between protein

aggregation and neurodegenerative disease. We focus on AD, because it set the historical course for the development of the field. AD is also the most prevalent human neurodegenerative disease, and, for that reason, is the subject of intense drug-development efforts. However, examples from other neurodegenerative diseases are discussed where appropriate. We also examine the state-of-the-art working hypotheses, asking what can be done to refine and clarify the current models, and more importantly, what needs to be done in order to produce disease-modifying treatments for neurodegenerative diseases and to identify presymptomatic individuals who could benefit from such treatments.

Fibrillar aggregates could be the cause or effect of disease

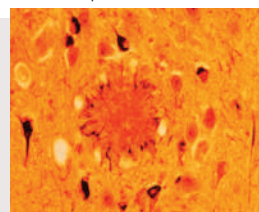
The two early hypotheses (Fig. 1) in this field were predicated on the fact that fibrillar deposits were the only observable proteinaceous abnormalities in postmortem human tissue. The two are mutually exclusive, but do not, as explained below, encompass all possibilities: both could be wrong. Although they differ in terms of the causal relationship between protein fibrils and disease, both posit that fibrillar aggregates and disease are linked.

As the list of neurodegenerative diseases characterized by fibrillar aggregates grew¹⁴, it became apparent that subtypes of several such diseases are not characterized by the presence of fibrillar deposits in the brain. This finding indicates that fibrillar deposits are not generally coupled to disease and, therefore, that hypotheses one and two are both incorrect. It is possible that diseases that do not show fibrillar aggregation, which include Gerstmann–Sträussler syndrome (a genetic prion disease)¹⁵, autosomal recessive juvenile parkinsonism (AJPR, a juvenile-onset form of PD)¹⁶ and sporadic ALS¹⁷, are not representative of most neurodegenerative diseases. However, the number and size of plaques in a postmortem AD brain do not correlate with the severity of symptoms at the time of death^{18,19}, which also argues against hypotheses one and two. Finally, another fact that conflicts with the second hypothesis is that amyloid plaques are found throughout the cortex of many cognitively normal 70-year-olds²⁰.

Fibrillar aggregates are an epiphenomenon of disease

So, a revised hypothesis is necessary to explain the qualitative, but not quantitative, correlation between fibrillar deposits and disease. This model (hypothesis number three, Fig. 2) allows for the extreme situations described above — prevalent deposits accompanied by little or no disease on the one hand and severe disease with little or no deposits on the other.

Prevalent deposits but little or no disease



Severe disease with little or no deposits

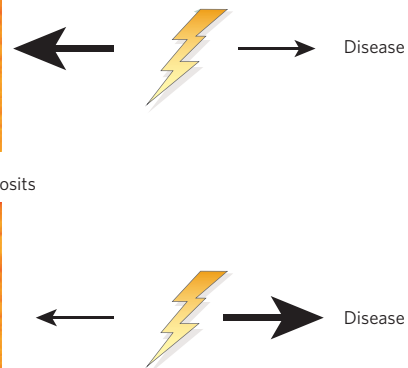
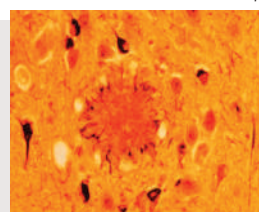


Figure 2 | Fibrillar deposits are not quantitatively correlated with disease. There is a qualitative, but not quantitative, correlation between fibrillar deposits and disease. One perturbation initiates both fibril formation and disease (lightning bolt), although the agent responsible for this is unknown.

According to this hypothesis, many different events (the lightning bolts in Fig. 2) could cause both fibril formation and disease. However, given that the fibrillar protein aggregates arise from a normally expressed protein (such as A β or α -synuclein) by a complex pathway, it is logical to start by postulating that a precursor to the fibrillar aggregates may be the triggering factor²¹. This hypothesis is supported by studies in animal models, which reinforce the conclusion that fibrillar deposits are not directly related to disease^{22–24}, and in some cases demonstrate that fibrillar deposits and disease-like phenotype are anticorrelated. Thus, perturbations that increase deposits decrease the disease-like phenotype and vice versa^{23,25,26}. The *in vitro* discovery and characterization of pre-fibrillar protein aggregates provided a framework for this hypothesis.

Protofibrils are early intermediates in fibril formation

Although conventional light microscopic analysis of postmortem AD brains revealed only fibrillar protein deposits, immunostaining with antibodies to A β allowed detection of another type of A β deposit, known as diffuse amyloid²⁷. This deposit lacks fibrillar substructure and was proposed to be a precursor to fibrillar amyloid. In contrast to the fibrillar deposits, which can be extracted from the brain because of their unusual stability, diffuse amyloid proved impossible to extract and characterize²⁸. Therefore, the first non-fibrillar A β aggregates to be characterized were produced *in vitro*, from synthetic A β .

Two groups, using complementary methods, demonstrated that A β produces non-fibrillar aggregates, which are consumed as fibrils are formed^{29,30}. These species were designated protofibrils, because the time dependence of their formation and the disappearance of their secondary structure (primarily β -sheet) were consistent with their being an intermediate in the fibrillization process (whether or not protofibrils are on the direct pathway from monomers to fibrils is still unclear). Notably, both groups reported that A β 42 forms protofibrils more rapidly than A β 40, which is consistent with the fact that A β 42 is much more strongly correlated with the disease process.

Subsequently, other non-fibrillar A β aggregates have been described, including A β -derived diffusible ligands (ADDLs)³¹, A β amylopherooids³², membrane-associated A β oligomers and annular A β protofibrils^{33,34}. How these species are related to one another is difficult to determine, because the conditions for their formation and characterization vary widely (Fig. 3). Furthermore, although it is likely that these metastable oligomeric species³⁵ assemble in a stepwise process involving

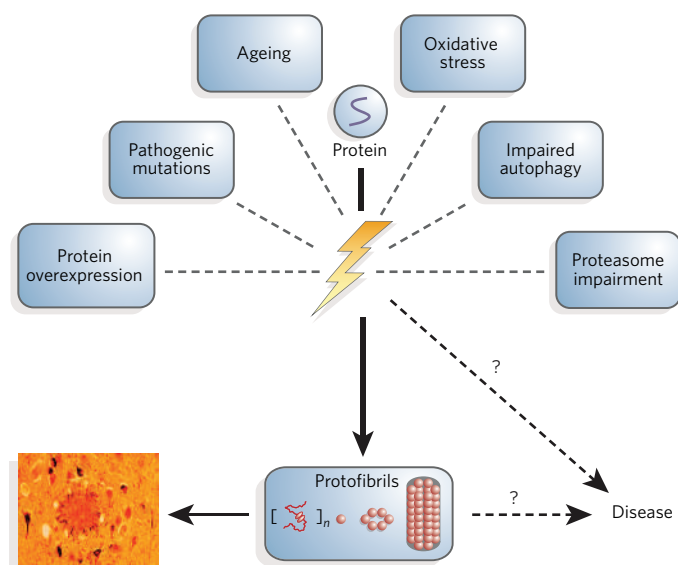
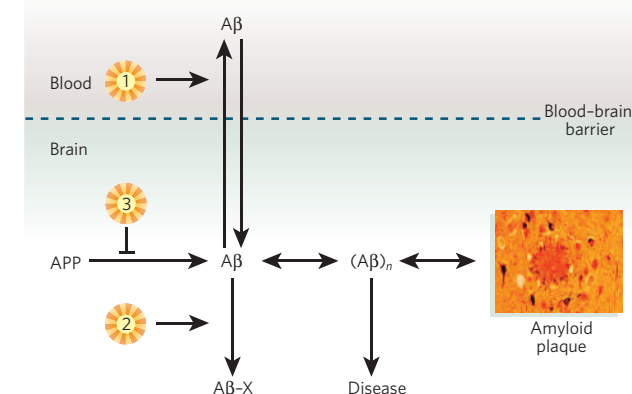


Figure 3 | Disease and fibril formation may have a common cause. A number of factors have been proposed to trigger protein oligomerization (protofibril formation) and disease. Whether protofibrillar aggregates are the cause of disease is uncertain, but circumstantial evidence supports a pathogenic role for these structures.

Box 2 | Clinical trials for strategies to reduce A β aggregation in AD



Three therapeutic strategies for Alzheimer's disease, all based on reducing the amount of protofibrillar A β aggregation ((A β)_n), have reached phase II clinical trials. Each involves a different mechanism (see panel).

The first of these is Elan's A β vaccine. Inoculation of mice with an A β -derived immunogen resulted in the reduction of amyloid plaques⁸² and a degree of rescue of some AD-like phenotypes^{24,83}. In subsequent animal studies, it was shown that passive immunization, by infusion of A β -directed antibodies, has a similar effect⁸⁴, leading to the hypothesis that the vaccine causes redistribution of A β from the brain into the periphery, where it can be degraded⁸⁵ (1). A reduction in brain A β would slow or even stop new aggregation, leading to a decrease in the rate of neurodegeneration. In addition, such a reduction could lead to complete resolution of amyloid plaques, which seem to be in dynamic equilibrium with monomeric A β and thus may be actively cleared.

A second strategy (2) involves the inhibition of all A β aggregation by binding monomeric A β with a small molecule (X; Neurochem's 'Alzhmed' (tramiprosate)) that is thought to be a proteoglycan mimic. This reduces amyloid-plaque formation in mouse models (and also rescues certain AD-like phenotypes)⁸⁶. The compound is well-tolerated and can cross the blood-brain barrier. Phase II trials suggested that a subset of patients with mild AD experienced some improvement in cognitive measures after receiving treatment for more than 1 year (<http://www.neurochem.com/ResearchActivities.htm>). Phase III trials, the ultimate determinant of efficacy, are underway.

A third strategy (3) involves the reduction of A β 42 production from APP (and A β 42 aggregation), by modulating γ -secretase/presenilin activity. This strategy originated with the epidemiological finding that chronic arthritis patients had a lower incidence of AD⁸⁷. It was determined that the use of some, although not all, non-steroidal anti-inflammatory drugs (NSAIDs) was responsible for this effect, and that the efficacy of NSAIDs did not correlate with their ability to inhibit cyclooxygenase 2 (COX2)⁸⁸. Myriad Genetics (<http://www.myriad.com>) is developing a compound, MPC-7869 (*R*-flurbiprofen, known as 'Flurizan'), that is one component of the racemic drug Flurbiprofen (Ansaid) and is itself inactive with respect to COX2 inhibition. MPC-7869 has been shown to reduce the levels of A β 42 in cell culture⁸⁹ and in animal models of AD⁹⁰ by modulating the activity of γ -secretase, and is currently being evaluated in phase III clinical trials. It is hoped that this compound will not exhibit the side-effect profile associated with COX2 inhibitors. In a phase II study, patients with mild AD showed a decreased rate of decline in several clinical measures of AD severity, although moderately affected patients showed no significant improvement. Several pharmaceutical companies have reported that γ -secretase inhibitors, which lower A β 42 and A β 40 by an active-site-directed mechanism, are entering phase II trials.

An alternative strategy to reduce aggregate levels involves the stimulation of aggregate degradation (possibly by autophagy). Finally, the idea that the promotion of fibril formation (at the expense, presumably, of protofibril) would be therapeutic is based on the aggresome proposal of Kopito⁹¹ and has shown some merit in a *Drosophila* model of PD²⁵ and in cell culture. However, this strategy carries considerable risk, because fibrils themselves may be toxic⁷¹.

low-molecular-weight oligomers, the relative importance of these is difficult to assess because they are too unstable to characterize^{36–38}.

The intermediacy of stable, structured oligomers is a common feature of *in vitro* fibrillization by many disease-associated proteins, including α -synuclein^{39,40}, huntingtin⁴¹ and islet amyloid polypeptide^{42,43}, as well as by fibril-forming proteins that are unrelated to disease⁴⁴. Non-fibrillar aggregates have also been isolated from cell culture^{45,46}, from the brains of animal models of AD^{21,47} and from diseased human brains^{48–51}. Annular pore-like aggregates have been observed during the *in vitro* fibrillogenesis of many well-characterized amyloid-forming proteins⁵². Annular α -synuclein-containing structures, similar in morphology to the structures formed *in vitro*, have been isolated from postmortem brain tissue of patients with multiple system atrophy (MSA) (ref. 51 and H.A.L., unpublished observations). MSA, like Parkinson's disease and diffuse Lewy-body disease, is characterized by α -synuclein deposition in the form of cytoplasmic Lewy bodies.

Fibrillar aggregates and disease may have a common cause

A revised model of pathogenesis, which takes into account the complexity of the fibrillization process, states that one perturbation, or a combination of several (Fig. 3, lightning bolt), causes protofibril formation, fibril formation and disease. Whether fibril formation is caused by protofibril formation, as shown in Fig. 3, or merely co-occurs with protofibril formation is not as important, from a drug development perspective, as defining the relationship between protofibril formation and disease (dashed arrows). It is tempting to speculate that protofibril formation may cause both fibril formation and disease, but first, the possibility that protofibril formation and disease are caused by elevated levels of monomer protein must be considered (hypothesis number four; Fig. 3).

Accumulation of protein monomers is not the cause

Neurodegeneration and the expression levels of certain disease-associated proteins are clearly related. For example, triplication⁵³ or duplication⁵⁴ of the α -synuclein gene causes PD, and overexpression to a lesser extent due to a promoter polymorphism increases an individual's risk of PD⁵⁵. This dependence is consistent with aggregation being important for disease, but is it possible that the monomeric protein independently causes both aggregation and disease (Fig. 3, hypothesis number four)?

Several facts indicate (but do not prove) that the monomeric proteins are non-toxic. First, the disease-linked proteins — such as α -synuclein and superoxide dismutase (SOD1, known to be involved in ALS) — are highly expressed in the normal brain. Second, the disease-linked proteins are not homologous at the level of primary sequence or native tertiary structure, making it difficult to imagine that this diverse group of monomeric proteins activates a common toxic mechanism. However, all of the aggregates are characterized by β -sheet secondary structure that may be responsible for shared mechanisms of toxicity. Third, many of the mutations that cause familial neurodegenerative diseases do not significantly alter the encoded protein's native structure, but reduce its stability and hence its steady-state level *in vivo* while accelerating its aggregation *in vitro* (this is well documented for AD⁵⁶, PD^{39,57–59} and ALS⁶⁰). For example, the level of mutant SOD1 in the blood of patients with familial ALS is considerably lower than the level of wild-type SOD1, which is inconsistent with the monomeric protein's being toxic (furthermore, the amount of mutant protein is inversely correlated with the aggressiveness of disease)⁶¹. Fourth, cell death in neurodegenerative diseases exhibits stochastic behaviour and seems to be highly dependent on the level of protein — that is, small changes in expression level have a notable effect on disease onset and progression⁶². This nonlinear dependence is not consistent with the protein monomer's being the pathogenic species, but is easily explained by a protein-aggregation mechanism.

Finally, three features of monogenic neurodegenerative diseases are inconsistent with the fourth hypothesis. First, the age of symptom onset in Huntington's disease does not depend on the dosage of the abnormal gene (in other words whether it is heterozygous or homozygous)⁶³. This finding is inconsistent with hypothesis number four, but can be explained

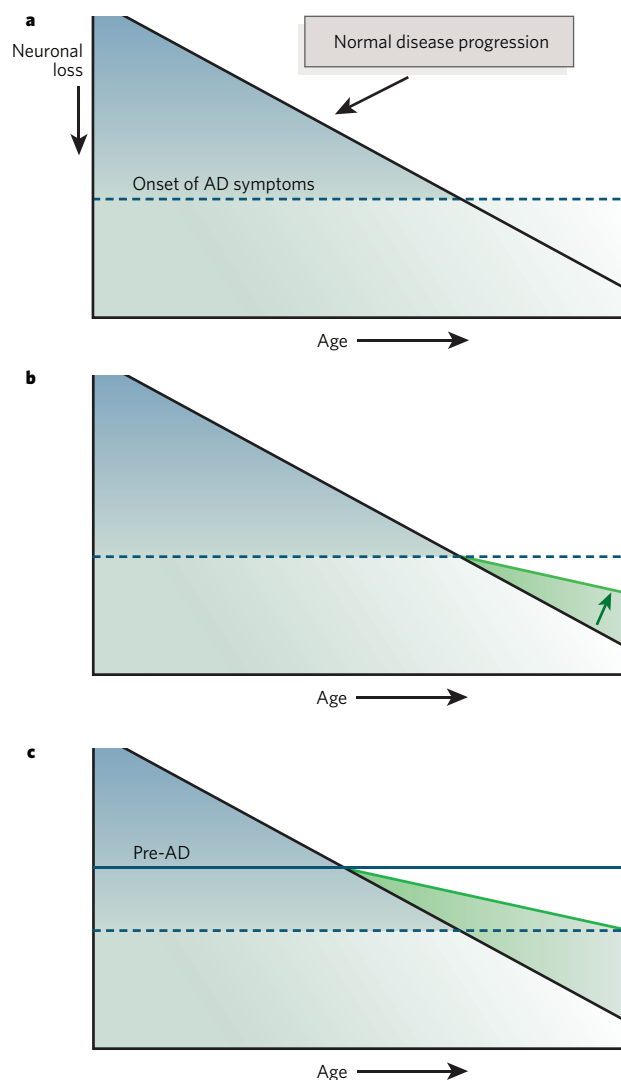


Figure 4 | Better diagnosis and presymptomatic treatment would drastically reduce the incidence and prevalence of neurodegenerative diseases. **a**, A diagram showing the loss of functional neurons over time that is characteristic of AD. Clinical diagnosis occurs when the disease is well underway, as symptoms do not become obvious until considerable neurodegeneration has occurred. **b**, A drug that slowed disease progression/neuronal loss, if started at clinical diagnosis, would have a significant effect on the impact of disease. **c**, If the same drug could be administered presymptotically to patients diagnosed by, for example, imaging methods, the disease might not progress to the point at which symptoms become obvious.

by a seeded-polymerization model in which aggregation is nucleated by the product of the abnormal gene, and the aggregate can then grow equally well with either the mutant or the wild-type protein⁶⁴.

Second, although an elevated level of plasma A β 42 is a risk factor for sporadic AD⁶⁵, several presenilin mutations linked to familial AD do not affect A β 42 levels and decrease A β 40 production (however, most familial-AD-linked presenilin mutations increase A β 42)⁶⁶. This effect is inconsistent with any hypothesis contending that A β 42 and/or A β 40 elevation alone causes familial AD.

Third, the fourth hypothesis is difficult to reconcile with a phenomenon observed in sporadic prion disease⁶⁷ and familial ALS^{68,69}, whereby mutations or polymorphisms on separate alleles of the gene encoding the aggregating protein (the prion protein and SOD1, respectively) modulate the resultant disease. This conflicts with the idea of a protein monomer's being responsible for neurotoxicity, in which case the effects of two mutations should be additive.

All of the data discussed above are consistent with the idea that non-fibrillar protein aggregates cause disease (hypothesis number five; Fig. 3). This hypothesis explains the lack of a gene-dosage effect in HD. It also explains why lowering A β 40 can cause familial AD — A β 40 inhibits the fibrillization of A β 42 (by stabilizing protofibrils) — and why raising the ratio of A β 42 to A β 40, whether by elevating A β 42 or by lowering A β 40, increases the rate of fibrillization⁷⁰. Finally, the phenomenon of compound heterozygosity is consistent with the fifth hypothesis, because intermolecular interactions are critical for determining aggregation rate and disease onset and progression.

Protofibrillar aggregates may trigger disease

A demonstration that clinical status and the abundance of certain prefibrillar protein aggregates are not quantitatively correlated would disprove the fifth hypothesis. However, at present this experiment is not possible, because there are no available imaging methods capable of tracking the early stages of protein aggregation *in vivo* (Box 1). This hypothesis could also be disproved if a drug that could selectively inhibit A β oligomerization and fibrillogenesis did not slow disease progression. In fact, three very different therapeutic strategies for AD that each seek to inhibit A β aggregation have shown some promise in reducing AD symptoms in phase II clinical trials (Box 2). It is important to note that the fifth hypothesis suggests that if fibrils are non-toxic (this is still an open question⁷¹), compounds that promote the conversion of early aggregates into fibrillar aggregates could slow disease progression.

Several protofibrillar aggregates may be neurotoxic

Because it is not currently possible to directly correlate the amount of a discrete protein aggregate to a human disease (Box 1), efforts have focused on cellular and mouse models of disease. Discrete protein aggregates have been identified both by *in vitro* aggregation of disease-associated proteins and in materials extracted from the brains of mouse models⁴⁷ and human patients^{50,51}. It is important to note that, in all of these, the extraction procedures and analysis methods (sodium dodecyl sulphate polyacrylamide gel electrophoresis, SDS-PAGE) are known to influence the dynamic aggregation process^{72,73}. Thus, it is never clear whether the species being manipulated in the laboratory ever existed in the tissue of interest (in theory, imaging could clarify this; see Box 1). However, bearing this caveat in mind, several compelling correlations have recently been reported. In each of these, the monomeric protein does not affect the phenotype of interest (providing further evidence against the second hypothesis) and in many (but not all⁷¹) cases, the fibrillar aggregate also has no effect.

Non-fibrillar A β preparations enriched in SDS-stable oligomers, spherical 4–6-nm aggregates and chain-like aggregates have all been shown to be toxic to cultured neurons, to inhibit hippocampal long-term potentiation (LTP)^{31,32,36,74,75}, to impair synaptic functions and to disrupt cognition and learned behaviour in rats^{47,76}. Non-fibrillar α -synuclein aggregates inhibit proteasomal activity⁷⁷ and induce Golgi fragmentation⁷⁸. Furthermore, non-fibrillar aggregate preparations produced from several amyloid-forming proteins, regardless of differences in their size and/or morphological distribution, have been shown to be toxic to cultured cells/neurons, suggesting that a wide range of different aggregate species may be neurotoxic.

The problem with all of the correlative studies done so far is that they each reduce a complex human disease to a single measurable phenomenon — for example, ‘the memory dysfunction in AD results from LTP deficit’. This oversimplification is required in order to reach clear conclusions, but AD (and PD, ALS and other similar disorders) probably results from a combination of biochemical, cell biological and systemic events (such as LTP deficit) each of which may, in turn, be influenced by one or more aggregate species. Ultimately, postmortem pathological data from human clinical trials of drug candidates that target specific processes (Box 2) will clarify the situation. However, in order to obtain this data, we must recognize that our current level of understanding, although incomplete, is sufficient for the development of such drug candidates.

We can change the medical landscape

It is critical to recognize that the onset of the neurodegenerative-disease process — which proceeds from the start of neuronal dysfunction — typically predates the first recognition of symptoms by at least 5–10 years (Box 1). Thus, if the presymptomatic disease could be diagnosed and treated, the prevalence of symptomatic disease could be greatly reduced (Fig. 4). Because most neurodegenerative diseases are genetically complex (HD, which is purely familial and monogenic, is the exception), presymptomatic diagnosis will require either the discovery of a peripheral marker for neurodegeneration that reacts earlier and is more sensitive than the traditional symptoms (for example, alteration of gene expression in the periphery), or that the process of protein aggregation be observable in the brain (Box 1). The development of useful imaging technology does not require that the relationship between aggregation and disease be explained in detail, only that a prefibrillar target be identified and that a target-specific probe exist. Once early protein aggregation has been detected, the patient will be treated with a disease-modifying drug designed to reduce the amount of aggregates in the brain. Once again, it would not be necessary to identify one specific target providing that the strategy used reduced all aggregation — for example, by reducing production of the aggregating protein (Box 2) or by stimulating aggregate clearance^{38,79,80}. Ultimately, Alzheimer's 100-year-old question, “what is the relationship between protein aggregation and neurodegeneration?” cannot be answered until the effects of such compounds on human disease are evaluated. The clinical trials will be large, time-consuming, risky and expensive, but we cannot afford to wait. ■

1. Alzheimer, A. Über eine eigenartige Erkrankung der Hirnrinde. *Allg. Z. Psychiatr. Psychiatr. Gerichl. Med.* **64**, 146–148 (1907).
2. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706 (1991).
3. Glenner, G. G. & Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120**, 885–890 (1984).
4. Masters, C. L. *et al.* Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl Acad. Sci. USA* **82**, 4245–4249 (1985).
5. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
6. Klunk, W. E. *et al.* Binding of the positron emission tomography tracer Pittsburgh compound-B reflects the amount of amyloid- β in Alzheimer's disease brain but not in transgenic mouse brain. *J. Neurosci.* **25**, 10598–10606 (2005).
7. Mathis, C. A., Klunk, W. E., Price, J. C. & DeKosky, S. T. Imaging technology for neurodegenerative diseases: progress toward detection of specific pathologies. *Arch. Neurol.* **62**, 196–200 (2005).
8. Fagan, A. M. *et al.* Inverse relation between *in vivo* amyloid imaging load and cerebrospinal fluid A β 42 in humans. *Ann. Neurol.* **59**, 512–519 (2006).
9. Dodart, J. C., Mathis, C., Bales, K. R. & Paul, S. M. Does my mouse have Alzheimer's disease? *Genes Brain Behav.* **1**, 142–155 (2002).
10. Fernagut, P. O. & Chesselet, M. F. Alpha-synuclein and transgenic mouse models. *Neurobiol. Dis.* **17**, 123–130 (2004).
11. Roses, A. D. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* **47**, 387–400 (1996).
12. Satoh, J. & Kuroda, Y. A polymorphic variation of serine to tyrosine at codon 18 in the ubiquitin C-terminal hydrolase-L1 gene is associated with a reduced risk of sporadic Parkinson's disease in a Japanese population. *J. Neurol. Sci.* **189**, 113–117 (2001).
13. Jarrett, J. T. & Lansbury, P. T. Seeding “one-dimensional crystallization” of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055–1058 (1993).
14. Ross, C. A. & Poirier, M. A. Opinion: what is the role of protein aggregation in neurodegeneration? *Nature Rev. Mol. Cell Biol.* **6**, 891–898 (2005).
15. Hsiao, K. & Prusiner, S. B. Molecular genetics and transgenic model of Gertsman-Straussler-Scheinker disease. *Alzheimer Dis. Assoc. Disord.* **5**, 155–162 (1991).
16. Kitada, T. *et al.* Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
17. Rosen, A. D. Amyotrophic lateral sclerosis. Clinical features and prognosis. *Arch. Neurol.* **35**, 638–642 (1978).
18. Terry, R. D. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* **30**, 572–580 (1991).
19. Lemere, C. A. *et al.* Sequence of deposition of heterogeneous amyloid β -peptides and APO E in Down syndrome: implications for initial events in amyloid plaque formation. *Neurobiol. Dis.* **3**, 16–32 (1996).
20. Dickson, D. W. *et al.* Identification of normal and pathological aging in prospectively studied nondemented elderly humans. *Neurobiol. Aging* **13**, 179–189 (1992).
21. Kuo, Y. M. *et al.* Comparative analysis of amyloid- β chemical structure and amyloid plaque morphology of transgenic mouse and Alzheimer's disease brains. *J. Biol. Chem.* **276**, 12991–12998 (2001).
22. Slow, E. J. *et al.* Absence of behavioral abnormalities and neurodegeneration *in vivo* despite widespread neuronal huntingtin inclusions. *Proc. Natl Acad. Sci. USA* **102**, 11402–11407 (2005).

23. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805–810 (2004).
24. Morgan, D. et al. A β peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature* **408**, 982–985 (2000).
25. Chen, L. & Feany, M. B. Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a *Drosophila* model of Parkinson disease. *Nature Neurosci.* **8**, 657–663 (2005).
26. Bowman, A. B., Yoo, S. Y., Dantuma, N. P. & Zoghbi, H. Y. Neuronal dysfunction in a polyglutamine disease model occurs in the absence of ubiquitin-proteasome system impairment and inversely correlates with the degree of nuclear inclusion formation. *Hum. Mol. Genet.* **14**, 679–691 (2005).
27. Nagy, Z. et al. The effects of additional pathology on the cognitive deficit in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* **56**, 165–170 (1997).
28. Roher, A. E. & Kuo, Y. M. Isolation of amyloid deposits from brain. *Methods Enzymol.* **309**, 58–67 (1999).
29. Harper, J. D., Wong, S. S., Lieber, C. M. & Lansbury, P. T. Observation of metastable A β amyloid protofibrils by atomic force microscopy. *Chem. Biol.* **4**, 119–125 (1997).
30. Walsh, D. M., Lomakin, A., Benedek, G. B., Condron, M. M. & Teplow, D. B. Amyloid β -protein fibrillogenesis. Detection of a protofibrillar intermediate. *J. Biol. Chem.* **272**, 22364–22372 (1997).
31. Lambert, M. P. et al. Diffusible, nonfibrillar ligands derived from A β 1–42 are potent central nervous system neurotoxins. *Proc. Natl Acad. Sci. USA* **95**, 6448–6453 (1998).
32. Hoshi, M. et al. Spherical aggregates of β -amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3 β . *Proc. Natl Acad. Sci. USA* **100**, 6370–6375 (2003).
33. Quist, A. et al. Amyloid ion channels: a common structural link for protein-misfolding disease. *Proc. Natl Acad. Sci. USA* **102**, 10427–10432 (2005).
34. Lashuel, H. A., Hartley, D., Petre, B. M., Walz, T. & Lansbury, P. T. Neurodegenerative diseases: amyloid pores from pathogenic mutations. *Nature* **418**, 291 (2002).
35. Lashuel, H. et al. Mixtures of wild-type and "Arctic" A β 40 *in vitro* accumulate protofibrils, including amyloid pores. *J. Mol. Biol.* **332**, 795–808 (2003).
36. Walsh, D. M. et al. Naturally secreted oligomers of amyloid β protein potentially inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416**, 535–539 (2002).
37. Townsend, M., Shankar, G. M., Mehta, T., Walsh, D. M. & Selkoe, D. J. Effects of secreted oligomers of amyloid β -protein on hippocampal synaptic plasticity: a potent role for trimers. *J. Physiol. (Lond.)* **572**, 477–492 (2006).
38. Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age onset proteotoxicity. *Science* advance online publication, 10 August 2006 (doi:10.1126/science.1124646).
39. Conway, K. A. et al. Acceleration of oligomerization, not fibrillization, is a shared property of both α -synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl Acad. Sci. USA* **97**, 571–576 (2000).
40. Rochet, J. C., Conway, K. A. & Lansbury, P. T. Inhibition of fibrillization and accumulation of prefibrillar oligomers in mixtures of human and mouse α -synuclein. *Biochemistry* **39**, 10619–10626 (2000).
41. Poirier, M. A. et al. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrillogenesis. *J. Biol. Chem.* **277**, 41032–41037 (2002).
42. Anguiano, M., Nowak, R. J. & Lansbury, P. T. Protofibrillar islet amyloid polypeptide permeabilizes synthetic vesicles by a pore-like mechanism that may be relevant to type II diabetes. *Biochemistry* **41**, 11338–11343 (2002).
43. Green, J. D., Goldsberry, C., Kistler, J., Cooper, G. S. & Aeby, U. Human amylin oligomer growth and fibril elongation define two distinct phases in amyloid formation. *J. Biol. Chem.* **279**, 12206–12212 (2004).
44. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
45. Mukai, H. et al. Formation of morphologically similar globular aggregates from diverse aggregation-prone proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **102**, 10887–10892 (2005).
46. Lee, H. J. & Lee, S. J. Characterization of cytoplasmic α -synuclein aggregates. Fibril formation is tightly linked to the inclusion-forming process in cells. *J. Biol. Chem.* **277**, 48976–48983 (2002).
47. Lesne, S. et al. A specific amyloid- β protein assembly in the brain impairs memory. *Nature* **440**, 352–357 (2006).
48. Riesner, D. et al. Disruption of prion rods generates 10-nm spherical particles having high alpha-helical content and lacking scrapie infectivity. *J. Virol.* **70**, 1714–1722 (1996).
49. Silveira, J. R. et al. The most infectious prion protein particles. *Nature* **437**, 257–261 (2005).
50. Roher, A. E. et al. Morphology and toxicity of A β -(1–42) dimer derived from neuritic and vascular amyloid deposits of Alzheimer's disease. *J. Biol. Chem.* **271**, 20631–20635 (1996).
51. Pountney, D. L. et al. Annular α -synuclein species from purified multiple system atrophy inclusions. *J. Neurochem.* **90**, 502–512 (2004).
52. Lashuel, H. A. & Lansbury, P. T. J. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q. Rev. Biophys.* advance online publication 18 September 2006 (doi:10.1017/S0033588306004422).
53. Singleton, A. B. et al. α -Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
54. Chartier-Harlin, M. C. et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169 (2004).
55. Maraganore, D. M. et al. Collaborative analysis of α -synuclein gene promoter variability and Parkinson disease. *J. Am. Med. Assoc.* **296**, 661–670 (2006).
56. Nilsberth, C. et al. The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A β protofibril formation. *Nature Neurosci.* **4**, 887–893 (2001).
57. Conway, K. A., Harper, J. D. & Lansbury, P. T. Accelerated *in vitro* fibril formation by a mutant α -synuclein linked to early-onset Parkinson disease. *Nature Med.* **4**, 1318–1320 (1998).
58. El-Agnaf, O. M. et al. Aggregates from mutant and wild-type α -synuclein proteins and NAC peptide induce apoptotic cell death in human neuroblastoma cells by formation of beta-sheet and amyloid-like filaments. *FEBS Lett.* **440**, 71–75 (1998).
59. Greenbaum, E. A. et al. The E46K mutation in α -synuclein increases amyloid fibril formation. *J. Biol. Chem.* **280**, 7800–7807 (2005).
60. Hart, P. J. Pathogenic superoxide dismutase structure, folding, aggregation and turnover. *Curr. Opin. Chem. Biol.* **10**, 131–138 (2006).
61. Sato, T. et al. Rapid disease progression correlates with instability of mutant SOD1 in familial ALS. *Neurology* **65**, 1954–1957 (2005).
62. Perutz, M. F. & Windle, A. H. Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats. *Nature* **412**, 143–144 (2001).
63. Gusella, J. F. & MacDonald, M. E. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nature Rev. Neurosci.* **1**, 109–115 (2000).
64. Lansbury, P. T. Structural neurology: are seeds at the root of neuronal degeneration? *Neuron* **19**, 1151–1154 (1997).
65. Mayeux, R. et al. Plasma amyloid β -peptide 1–42 and incipient Alzheimer's disease. *Ann. Neurol.* **46**, 412–416 (1999).
66. Cupers, P. et al. The discrepancy between presenilin subcellular localization and γ -secretase processing of amyloid precursor protein. *J. Cell Biol.* **154**, 731–740 (2001).
67. Pardi, P. et al. Molecular basis of phenotypic variability in sporadic Creutzfeldt-Jakob disease. *Ann. Neurol.* **39**, 767–778 (1996).
68. Hand, C. K. et al. Compound heterozygous D90A and D96N SOD1 mutations in a recessive amyotrophic lateral sclerosis family. *Ann. Neurol.* **49**, 267–271 (2001).
69. Restagno, G. et al. The IVS1 +319 t>a of SOD1 gene is not an ALS causing mutation. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **6**, 45–49 (2005).
70. Hasegawa, K., Yamaguchi, I., Omata, S., Gejyo, F. & Naiki, H. Interaction between A β (1–42) and A β (1–40) in Alzheimer's β -amyloid fibril formation *in vitro*. *Biochemistry* **38**, 15514–15521 (1999).
71. Isaacs, A. M., Senn, D. B., Yuan, M., Shine, J. P. & Yankner, B. A. Acceleration of amyloid β -peptide aggregation by physiological concentrations of calcium. *J. Biol. Chem.* advance online publication 26 July 2006 (doi:10.1074/jbc.M602061200).
72. Yu, W. et al. Oligomerization of amyloid β -protein occurs during the isolation of lipid rafts. *J. Neurosci. Res.* **80**, 114–119 (2005).
73. Bitan, G., Fradinger, E. A., Spring, S. M. & Teplow, D. B. Neurotoxic protein oligomers — what you see is not always what you get. *Amyloid* **12**, 88–95 (2005).
74. Hartley, D. M. et al. Protofibrillar intermediates of amyloid β -protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J. Neurosci.* **19**, 8876–8884 (1999).
75. White, J. A., Manelli, A. M., Holmberg, K. H., Van Eldik, L. J. & Ladu, M. J. Differential effects of oligomeric and fibrillar amyloid- β 1–42 on astrocyte-mediated inflammation. *Neurobiol. Dis.* **18**, 459–465 (2005).
76. Cleary, J. P. et al. Natural oligomers of the amyloid- β protein specifically disrupt cognitive function. *Nature Neurosci.* **8**, 79–84 (2005).
77. Lindersson, E. et al. Proteasomal inhibition by α -synuclein filaments and oligomers. *J. Biol. Chem.* **279**, 12924–12934 (2004).
78. Gosavi, N., Lee, H. J., Lee, J. S., Patel, S. & Lee, S. J. Golgi fragmentation occurs in the cells with prefibrillar α -synuclein aggregates and precedes the formation of fibrillar inclusion. *J. Biol. Chem.* **277**, 48984–48992 (2002).
79. Komatsu, M. et al. Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441**, 880–884 (2006).
80. Hara, T. et al. Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
81. Kruger, R. et al. Familial parkinsonism with synuclein pathology: clinical and PET studies of A30P mutation carriers. *Neurology* **56**, 1355–1362 (2001).
82. Schenk, D. et al. Immunization with amyloid- β attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
83. Janus, C. et al. A β peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature* **408**, 979–982 (2000).
84. Bard, F. et al. Peripherally administered antibodies against amyloid β -peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
85. DeMattos, R. B. et al. Peripheral anti-A β antibody alters CNS and plasma A β clearance and decreases brain A β burden in a mouse model of Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **98**, 8850–8855 (2001).
86. Gervais, F. et al. Targeting soluble A β peptide with Tramiprosate for the treatment of brain amyloidosis. *Neurobiol. Aging* (in the press) published online 3 May 2006 (doi:10.1016/j.neurobiolaging.2006.02.015).
87. McGeer, P. L., Schulzer, M. & McGeer, E. G. Arthritis and anti-inflammatory agents as possible protective factors for Alzheimer's disease: a review of 17 epidemiologic studies. *Neurology* **47**, 425–432 (1996).
88. Weggen, S. et al. A subset of NSAIDs lower amyloidogenic A β 42 independently of cyclooxygenase activity. *Nature* **414**, 212–216 (2001).
89. Morihiro, T., Chu, T., Ubeda, O., Beech, W. & Cole, G. M. Selective inhibition of A β 42 production by NSAID R-enantiomers. *J. Neurochem.* **83**, 1009–1012 (2002).
90. Eriksen, J. L. et al. NSAIDs and enantiomers of flurbiprofen target γ -secretase and lower A β 42 *in vivo*. *J. Clin. Invest.* **112**, 440–449 (2003).
91. Kopito, R. R. Aggregates, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524–530 (2000).

Acknowledgements P.T.L. thanks the National Institutes of Health for supporting his work in this area. H.A.L. thanks R. Luthi-Carter, P. Fraering and N. Zavlia for reviewing the manuscript, and the Swiss National Science Foundation for supporting research in his laboratory.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to P.T.L. (lansbury@rics.bwh.harvard.edu).

The roles of intracellular protein-degradation pathways in neurodegeneration

David C. Rubinsztein¹

Many late-onset neurodegenerative diseases, including Parkinson's disease and Huntington's disease, are associated with the formation of intracellular aggregates by toxic proteins. It is therefore crucial to understand the factors that regulate the steady-state levels of these 'toxins', at both the synthetic and degradation stages. The degradation pathways acting on such aggregate-prone cytosolic proteins include the ubiquitin-proteasome system and macroautophagy. Dysfunction of the ubiquitin-proteasome or macroautophagy pathways might contribute to the pathology of various neurodegenerative conditions. However, enhancing macroautophagy with drugs such as rapamycin could offer a tractable therapeutic strategy for a number of these diseases.

Many late-onset neurodegenerative diseases are caused by aggregate-prone proteins¹. These diseases, known as proteinopathies, include conditions in which the proteins are predominantly cytosolic (such as Parkinson's disease, PD, and adult-onset Huntington's disease, HD), predominantly intranuclear (for example, spinocerebellar ataxia type 1), aggregated in the endoplasmic reticulum (as seen with neuroserpin mutations that cause familial encephalopathy with neuroserpin inclusion bodies) and secreted extracellularly (for example, amyloid- β in Alzheimer's disease). Whether soluble monomers, oligomers or larger aggregates are the most toxic species causing many of these conditions has been the subject of considerable debate and uncertainty (for a discussion, see ref. 2 and page 774). Nevertheless, in general, it seems that the capacity to aggregate (although not necessarily the aggregates themselves) is correlated with toxicity. For instance, the point mutations in α -synuclein and tau that cause dominant forms of PD and frontotemporal dementia, respectively, aggregate more rapidly than the wild-type proteins. The same phenomenon is seen with the polyglutamine expansion mutations that cause diseases such as HD — in these diseases earlier onset correlates with longer polyglutamine tracts, and aggregation propensity increases with polyglutamine tract length³.

Extensive genetic and transgenic data suggest that many of the mutations responsible for proteinopathies cause disease by conferring a toxic gain of function on the relevant proteins (for example, HD; see Box 1). Thus, it is important to understand the cellular processes regulating their levels. Steady-state levels are determined by synthesis and degradation rates.

For some such proteins, it is likely that the holoprotein is relatively or entirely benign, and that the toxic species is generated by a series of cleavage events leading to the production of a toxic fragment. In such cases, the rate of synthesis of the full-length protein may not be a major determinant of the production rate of the toxic species — the main rate-limiting step under conditions of synthesis varying between 50% and 100% of physiological rates may be at the level of cleavage (if the cleavage step is saturated). The best-known example of cleavage generating a toxic product occurs during the production of amyloid- β peptide from full-length amyloid precursor protein (APP)⁴. However, related toxic-fragment models also seem to be relevant to certain polyglutamine diseases,

such as HD^{5,6} and spinocerebellar ataxia type 3 (ref. 7). In principle, these processing steps may be therapeutically tractable. For instance, it may be possible to modulate amyloid- β levels by altering the activities of the different secretase enzymes that regulate its formation⁴.

In this review, I focus on how cells regulate the levels of aggregate-prone proteins, concentrating on degradation pathways. I consider the possibility that defective degradation causes certain forms of neurodegeneration, and discuss therapeutic strategies aimed at reducing the levels of toxic proteins by enhancing their degradation.

Pathways that clear aggregate-prone proteins

The ubiquitin-proteasome and autophagy-lysosome pathways are the two main routes of protein and organelle clearance in eukaryotic cells. Proteasomes are barrel-shaped multiprotein complexes that predominantly degrade short-lived nuclear and cytosolic proteins⁸ (Fig. 1). The ubiquitin-proteasome system is also important for the degradation of misfolded proteins in the endoplasmic reticulum. Misfolded endoplasmic-reticulum proteins are retrotranslocated back into the cytosol, where they are degraded by proteasomes (a process known as endoplasmic-reticulum-associated degradation, or ERAD). Most proteins are targeted for proteasomal degradation after being covalently modified with ubiquitin, which is conjugated through its carboxy terminus, usually to ϵ -amino groups of lysine residues. This conjugation typically involves three types of enzyme: E1 (ubiquitin-activating enzyme) hydrolyses ATP and forms a thioester-linked conjugate between itself and ubiquitin; E2 (ubiquitin-conjugating enzyme) receives ubiquitin from E1 and forms a similar thioester intermediate with ubiquitin; and E3 (ubiquitin ligase) binds both E2 and the substrate, and transfers the ubiquitin to the substrate. Frequently, the ubiquitin itself forms a substrate for further rounds of ubiquitination, resulting in the formation of a polyubiquitin chain. Chains of four or more ubiquitin molecules (linked by lysines at residue 48) appear to form a recognition signal that allows substrates to be shuttled to the proteasome by a set of proteins that, in at least some cases, include the chaperone CDC48/p97 (refs 9, 10).

Many of the proteins that cause proteinopathies (including those with polyglutamine expansions and α -synucleins) are partly dependent on the ubiquitin-proteasome pathway for their clearance^{11,12}.

¹Department of Medical Genetics, Cambridge Institute for Medical Research, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK.

Box 1 | Huntington's disease results from a gain-of-function mutation

Loss-of-function mutations cause disease by reducing the activity of the gene product. In an autosomal-dominant disease, the most severe mutations — null mutations — reduce protein activity to 50% of wild-type levels. The results of genetic and transgenic experiments strongly indicate that HD and related polyglutamine diseases are not caused by a simple hemizygous loss of function^{70–72}.

Mice lacking one HD-gene orthologue are essentially normal, although complete loss of huntingtin causes embryonic lethality. Humans with Wolf–Hirschhorn syndrome have hemizygous loss of the tip of chromosome 4p, which includes the HD gene. These people do not show features of HD⁷³. However, a number of transgenic mouse and *Drosophila* models have been made in which a mutant-huntingtin transgene, comprising either the whole coding region, an amino-terminal fragment or simply isolated expanded polyglutamine repeats, is expressed. These animals also have two normal huntingtin orthologues but recapitulate many of the features of the human disease (for a review, see ref. 74).

Furthermore, expression of hypoxanthine phosphoribosyl transferase with expanded polyglutamines caused the mice to develop a progressive neurological disorder with clinical and pathological features reminiscent of HD⁷⁵. Because complete loss of hypoxanthine phosphoribosyl transferase results in no overt phenotype in mice, this suggests that the predominant action of the HD mutation is a toxic gain of function. Gain-of-function mutations can cause disease by conferring additional properties on the mutant gene product — these may include hyperactivity of normal function and/or new toxic properties unrelated to normal function. However, it is possible that the HD mutation may result in some loss of function that may make minor contributions to pathogenesis⁷⁶.

In a dominant-negative mutation, the (autosomal dominant) mutation reduces the activity of the mutant allele's gene product but also inhibits the function of the wild-type product. This can occur if the gene encodes components of oligomeric complexes, in which optimal activity requires all components to be functional. In dominant-negative mutations, the activity of the gene products encoded by the wild-type and mutant gene is generally <50% of wild type. We cannot exclude the possibility that there may be some dominant-negative effects that impact on HD, in addition to gain of function.

However, substrates need to be unfolded to pass through the narrow pore of the proteasome barrel, which precludes the clearance of oligomeric and aggregated proteins¹³. Recent data suggest that the proteasome's enzymatic machinery might not be able to cleave between successive glutamine residues^{14,15}. This has important implications for polyglutamine-expansion mutations, as it suggests that the proteasome removes the flanking sequences around the expansions. On exiting from the proteasome, such isolated polyglutamine stretches may be more toxic than the pre-proteasomal species with the flanking sequences. However, it is likely that the isolated polyglutamine peptides exiting the proteasome are rapidly degraded by as yet unidentified cytosolic peptidases, because there are a number of normal proteins with wild-type polyglutamine stretches of 20–35 repeats.

Mammalian lysosomes can degrade substrates such as protein complexes and organelles. The bulk degradation of cytoplasmic proteins or organelles is largely mediated by macroautophagy, a process generally referred to as autophagy¹⁶ (Box 2). This involves the formation of double-membrane-bounded structures known as autophagosomes, or autophagic vacuoles (AVs). These fuse with lysosomes to form autophagolysosomes, and their contents are then degraded by acidic lysosomal hydrolases. Autophagosomes are generated by the elongation of small membranous structures known as autophagosome precursors. However, the precise origin or origins of these structures have yet to be elucidated in mammalian cells. Macroautophagy can be induced under conditions of physiological stress such as starvation. Several protein kinases regulate macroautophagy, the best characterized of which is mammalian target of rapamycin (mTOR), which negatively regulates the pathway. Many of the proteins with key roles in autophagy are

designated Atg (in yeast)/ATG (in mammals), and a number of these are conserved from yeast to humans.

Polyglutamine-expansion mutations, such as those seen in mutant huntingtin and ataxin 3 (causing spinocerebellar ataxia type 3, SCA3), mutant forms of α -synuclein (which cause familial PD), and different forms of tau (including mutations causing frontotemporal dementia) are strongly dependent on the macroautophagy pathway for their clearance^{12,17–21}. Pharmacological and genetic interference in this pathway slows the clearance of these substrates and increases levels of aggregates (where these are seen) and toxicity (in models showing toxicity). Inhibition of macroautophagy by gene knockout in otherwise normal mice leads to the formation of ubiquitinated aggregates in various tissues²².

Our data suggest that the dependence of proteins on the macroautophagy pathway for their clearance correlates with their propensity to aggregate^{11,12,17,20,23}. For instance, inhibition of macroautophagy has much smaller effects (if any effect at all) on the clearance of wild-type huntingtin exon 1 fragments or wild-type α -synuclein than on the clearance of the mutant aggregate-prone species^{11,12,17,23}. This might be because the wild-type proteins are efficiently cleared by proteasomes. The ubiquitin–proteasome system is more efficient than basal levels of macroautophagy, so for proteins that have access to both pathways, proteasomes are the favoured and dominating clearance route. This preference for ubiquitin–proteasomal degradation may also relate to tagging and targeting mechanisms that contribute to the selectivity of this pathway. When a cytosolic protein is aggregate-prone and a poor proteasome substrate, macroautophagy becomes a major clearance route by default — under these circumstances, the macroautophagy route becomes more effective than the proteasome (Fig. 2). Indeed, when normal cultured neurons are treated with proteasome inhibitors, the proteins that form aggregates seem to be degraded by macroautophagy²⁴. Although this model (Fig. 2) suggests that macroautophagy is a default pathway that becomes increasingly important when an aggregate-prone substrate cannot be efficiently cleared by the proteasome, it is possible that there are also signals (as yet unknown) that preferentially target such proteins to autophagosomes. Macroautophagy — in which bulk cytoplasm is engulfed by autophagosomes — is generally considered to be a non-specific process in organisms from yeast to humans, but a specific uptake process is conceivable, as there are precedents for cargo selection in the macroautophagy-related cytoplasm-to-vacuole trafficking (Cvt) pathway in yeast¹⁶.

Another pathway for cytosolic protein clearance through lysosomes is chaperone-mediated autophagy (CMA), a pathway distinct from macroautophagy. All the CMA substrate proteins described so far are soluble cytosolic proteins containing a targeting motif biochemically related to the pentapeptide Lys–Phe–Glu–Arg–Gln (KFERQ)²⁵. This motif, which is present in ~30% of cytosolic proteins, is recognized by a cytosolic chaperone, heat-shock cognate 70 (HSC70). This transfers protein substrates to the lysosomal membrane, where, through binding to the receptor lysosome-associated membrane protein 2A (LAMP2A), they are translocated into the lysosomal lumen and degraded. Lysosomes can also directly engulf cytoplasm by invagination, protrusion and/or septation of the lysosomal limiting membrane, a process known as microautophagy. However, this process is poorly understood in mammalian cells.

Proteasome dysfunction leading to neurodegeneration

A number of proteinopathies have been associated with decreased proteasome activity in humans or in cell models. These include sporadic PD, in which levels of 20/26S proteasomes and proteasome activity are reduced in vulnerable regions, and HD^{26,27}. Decreased proteasome-pathway activity seems to be a feature of many cell models in which intracellular aggregates are formed²⁸. However, it is not clear whether these phenomena are a cause or late effects of disease. In some polyglutamine-disease mouse models, proteasome impairment is not seen at (early) symptomatic stages^{29,30}. Furthermore, proteasome subunits can be cleaved by caspases in cells undergoing apoptosis³¹, and it is possible that this process may explain some of the cell-model data.

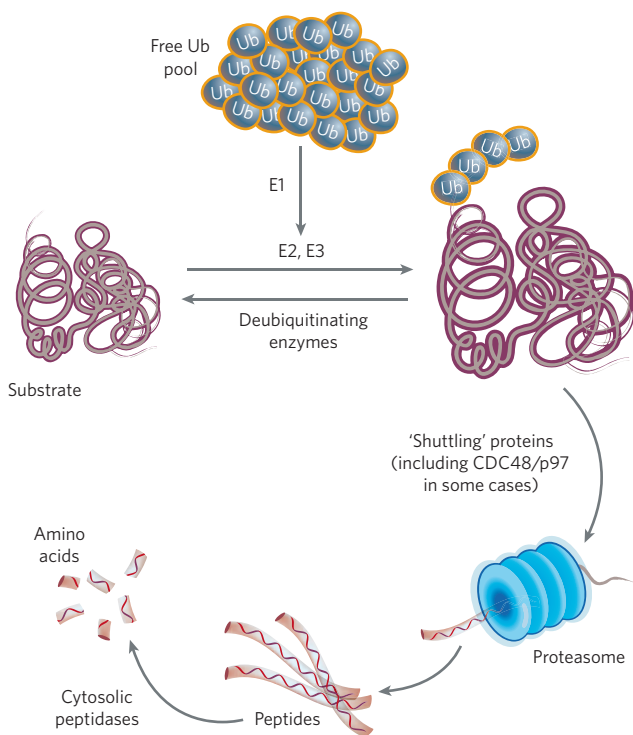


Figure 1 | Schematic diagram of the ubiquitin-proteasome system. Before they are targeted for proteasomal degradation, most proteins are covalently modified with ubiquitin (Ub). Typically, three enzyme types are involved in this process — ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin ligase (E3) enzymes. Proteins tagged with chains of four or more ubiquitins are shuttled to the proteasome by various proteins such as CDC48/p97. In the proteasome, proteins are reduced to peptides, which are then released into the cytosol and further broken down by peptidases. See the text for more details.

The strongest support for the possibility that the ubiquitin-proteasome system has a primary role in the pathology of a wide range of proteinopathies (or in sporadic forms of diseases such as PD) comes from human/mammalian mutations that involve components of this pathway and cause neurodegeneration. These include parkin, an E3 ligase in which autosomal recessive loss-of-function mutations cause PD^{32–34}. A number of parkin substrates have been identified, but at present it is not clear which is the major contributor to pathology in parkin deficiency. Heterozygous mutations in the gene encoding ubiquitin carboxy-terminal hydrolase L1 (UCHL1) have been causally implicated in families with PD³⁵, although the data are contentious³⁶. Although the S18Y variant has been associated with sporadic PD in genetic-association studies, this association has not withstood robust confirmatory analyses³⁷. UCHL1 might function not only as a ubiquitin hydrolase (hydrolysing ubiquitin chains to free ubiquitin monomers) but also as a ubiquitin ligase³⁸. The role of UCHL1 in human PD is still unclear, but loss of function is known to be deleterious, as recessive deletion of part of this gene causes gracile axonal dystrophy in mice³⁹. Although this phenotype is not obviously parkinsonian, it is associated with the formation of ubiquitinated intraneuronal aggregates³⁹.

Mutations in CDC48/p97 (also known as valosin-containing protein) cause a dominantly inherited disease known as inclusion-body myopathy with Paget's disease of bone and frontotemporal dementia, which is also characterized by cytoplasmic and nuclear aggregates in muscle and brain tissues⁴⁰. This pathology is consistent with the role of CDC48/p97 as a component of the machinery shuttling ubiquitinated ERAD substrates to the proteasome^{9,10} (although CDC48/p97 has a number of other roles that may also affect pathology).

The above genetic data, which suggest that primary genetic deficiencies of components of the ubiquitin-proteasome system are sufficient to

cause neurodegeneration, are complemented by pharmacological data showing that the injection of proteasome inhibitors into adult rats causes parkinsonian features, including Lewy-body-like aggregates⁴¹. However, it is still not clear whether proteasomal dysfunction is an important early causal factor *in vivo* in proteinopathies in which there are no primary defects in the ubiquitin-proteasome pathway.

Autophagy dysfunction and neurodegeneration

Macroautophagy (referred to below simply as autophagy) blockade — for example, through gene knockout — has deleterious effects on a number of systems. It prevents efficient nutrient recycling from macromolecules under conditions of starvation⁴², enhances susceptibility to certain types of apoptosis^{43,44} and, as mentioned previously, leads to the formation of ubiquitinated inclusions²⁴. Knockouts of specific autophagy genes (*Atg5* and *Atg7*) lead to early neonatal lethality^{22,42}. Interestingly, mice with neuronally confined autophagy-gene knockouts develop intraneuronal aggregates and neurodegeneration^{45,46}.

Although defects in autophagy-restricted genes have not been described in humans, autophagy may be compromised by other means. For example, dynein function is important for autophagosome-lysosome fusion, and autophagic clearance is impaired even by hemizygous loss of dynein activity⁴⁷. Dyneins are motor proteins that typically move cargoes towards the cell centre along microtubules, and components of this machinery are genetically compromised in some forms of motor neuron disease, either directly (for example, by a dynactin mutation in one form of human motor neuron disease^{48,49}, and dynein heavy-chain mutations in the legs at odd angles (*Loa*) and cramping (*Cra1*) mice)⁵⁰, or indirectly, if dynein functions are compromised by other mutations (for instance, SOD1 mutations causing familial amyotrophic lateral sclerosis^{51,52}). In some of these conditions (for example, SOD1 mutations, *Loa* mice and human dynactin mutations), protein aggregation is a feature, and this may be due in part to compromised autophagy. This possibility is supported by studies showing increased autophagosome numbers (compatible with decreased autophagosome-lysosome fusion) in the *Loa* mice, and increased aggregation and toxicity of mutant huntingtin fragments when HD mice are crossed with the *Loa* mice (or when the HD mutation is expressed in flies with hemizygous loss of either dynein heavy-chain or light-chain genes)⁴⁷. Although one cannot claim that defective autophagy is responsible for all of the pathology resulting from such mutations, by slowing the clearance of the aggregate-prone proteins and possibly increasing susceptibility to apoptosis, its contribution of dyneins is likely to be significant^{43,44}.

The chaperone-mediated autophagy pathway may also be relevant to neurodegeneration. Wild-type α -synuclein has recently been shown to be selectively translocated into lysosomes for degradation by the CMA pathway as well as being a substrate for the proteasome. In assays *in vitro* using isolated lysosomes, the pathogenic A53T and A30P α -synuclein mutant proteins were shown to bind to the CMA-pathway receptor on the lysosomal membrane, but seem to act as uptake blockers, inhibiting both their own degradation and that of other substrates⁵³. It will be interesting to investigate whether similar effects are mediated by these mutations *in vivo*, because such a mechanism might contribute to their toxic gain of function.

Therapeutic strategies to reduce toxin synthesis

Because many aggregate-prone proteins cause disease by gain-of-function mechanisms, efforts have been made to explore therapeutic strategies that reduce the levels of such proteins, at either the synthesis or degradation stage. Promising progress has been made towards selective synthesis inhibition of dominant gain-of-function mutations, such as the mutant HD allele⁵⁴, by RNA-interference (RNAi)-related approaches, as these strategies can discriminate targets that differ by a single nucleotide⁵⁵. Although this is theoretically attractive, there are a number of practical issues that still need to be resolved. First, safe, effective, long-term gene expression in the brain driven by viruses (or other possible vectors) remains a considerable challenge for gene therapy. Second, long-term RNAi may affect pathways unrelated to the

gene of interest. Third, if one is seeking to switch off the mutant allele in a polyglutamine disease, such as HD, without affecting the expression of the wild-type allele, this will probably require identification of a single-nucleotide polymorphism in the mutant allele that is different from the wild-type allele (as it may be impossible to specifically target the mutant allele on the basis of recognition of the expanded CAG repeats). Such polymorphisms will differ between patients, so a number of knockdown vectors and possibly even custom-designed gene therapy will be required. The latter option may not be feasible, given the safety testing that such procedures may require. However, I believe that there are major theoretical advantages to the RNAi-type approach that justify careful experimental assessment of its feasibility, as this strategy could have relevance to many diseases caused by gain-of-function mutations.

Enhancing autophagy as a therapeutic strategy

Knowledge of the degradation pathways acting on the toxic proteins that cause various proteinopathies could help in the design of therapeutic strategies. Upregulating proteasome activity may not be desirable for a number of reasons. The proteasome degrades many key short-lived

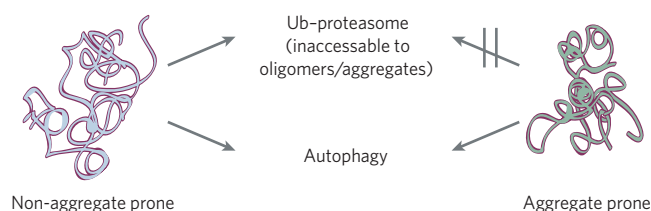


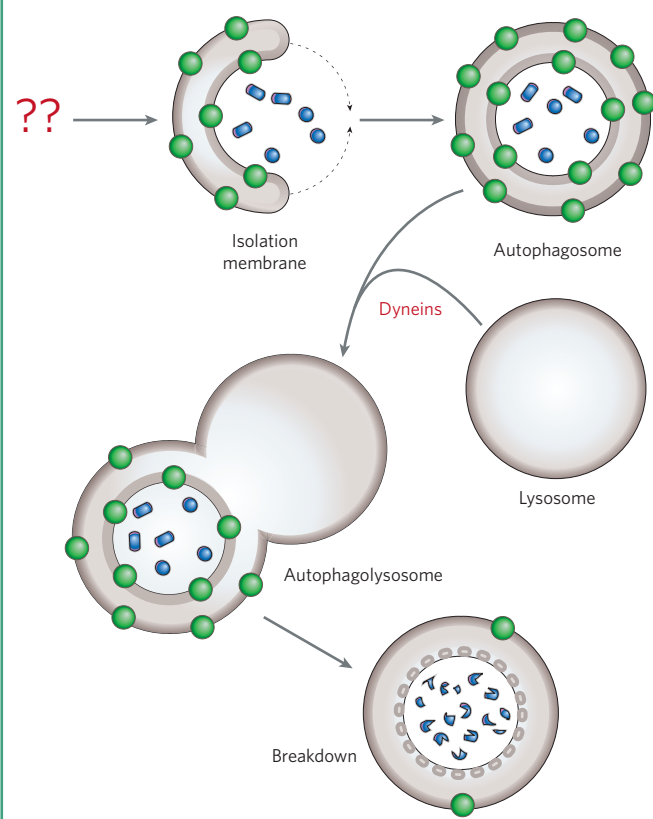
Figure 2 | Macroautophagy as a default pathway for proteasome-inaccessible substrates. When proteins are accessible to both the ubiquitin (Ub)–proteasome and autophagy pathways, the greater efficiency of the ubiquitin–proteasome system makes it the favoured and dominant clearance route. When a cytosolic protein is aggregate prone and a poor proteasome substrate, then autophagy becomes the main clearance route by default — under these circumstances, the autophagy route becomes more effective than the proteasome.

Box 2 | Schematic overview of key aspects of autophagy

Autophagy begins with the formation of double-membrane-bounded autophagosomes. The origin(s) of the autophagosome membranes are unclear. The mammalian target of rapamycin (mTOR) is a negative regulator of autophagosome formation, although how mTOR regulates this process in mammals is not clear. When mTOR is inhibited by rapamycin, autophagy is stimulated.

Autophagosomes fuse with lysosomes to form autophagolysosomes, a process that is governed by a number of factors, including dynein activity. The contents of autophagolysosomes are finally degraded by acidic lysosomal hydrolases.

The green dots represent the protein LC3 (also known as ATG8). This is the only known marker that specifically localizes to autophagosome and autophagolysosome membranes and not to other membranes. LC3 localizes to these structures after it has been processed and conjugated to phosphatidylethanolamine.



intracellular regulators (for example, p53), whose steady-state levels are dependent on their degradation rates, which can be selectively modulated by signals and post-translational modifications that influence processes such as ubiquitination. Enhancing the degradation of such proteins may have other deleterious consequences (such as cancer, in the case of p53). However, one way that proteasome clearance of aggregate-prone proteins might be enhanced without a corresponding effect on general ubiquitin–proteasome functioning could be to treat cells with chemical or molecular chaperones. By assisting normal folding and reducing aggregation of such proteins, chaperones shift the balance towards monomeric soluble forms (Fig. 2). These forms can often be cleared by the proteasome. The theoretical prediction that chaperone treatments might enhance proteasome-mediated clearance of such proteins and thereby reduce levels of the substrate is consistent with some experimental data^{56,57}. This strategy may be particularly valuable for clearing intranuclear proteins that are not accessible to autophagy.

Upregulating autophagy may be tractable and less toxic, as autophagy substrates are typically long-lived proteins and are not believed to be selectively degraded, in contrast to many ubiquitin–proteasome substrates. Less marked changes would be expected in steady-state levels of long-lived proteins compared with proteins with short half-lives if their clearance were perturbed to similar extents and for equivalent times. Upregulation of autophagy with the mTOR inhibitor rapamycin reduced levels of mutant soluble huntingtin (and aggregates) and attenuated mutant-huntingtin-fragment toxicity in cells, and in transgenic *Drosophila* and mouse models^{11,21}. We have confirmed that the main effect of rapamycin in this context occurs through autophagy²⁰. Subsequent studies have provided robust support for our assertions and suggest that autophagy is important even for the clearance of full-length mutant huntingtin, although it is not limiting for the clearance of its wild-type counterpart^{17–19}.

Recent data from studies of cell and fly models show that autophagy upregulation might be valuable for reducing toxicity in many intracellular proteinopathies other than HD, including those caused by polyglutamine expansions such as those seen in SCA3, mutant forms of α -synuclein, and both mutant and wild-type tau²⁰. Thus, autophagy upregulation may be beneficial in other polyglutamine diseases, Parkinson's disease, and some forms of frontotemporal dementia.

Rapamycin treatment of intracytosolic proteinopathies reduces the levels of soluble and aggregated proteins, in both cell and animal models, in an autophagy-dependent manner²⁰. However, it is not clear whether autophagy clears only soluble species and oligomers, or whether it can also directly clear larger aggregates. Our data — from mutant forms of α -synuclein that do not aggregate in the cell lines we used — suggest that this pathway can efficiently target species that are not in aggregates large enough to be seen with light microscopy¹². The observation of inclusion formation in the neurons of mice with neuronally restricted autophagy-gene knockouts is consistent with the ability of autophagy to clear soluble and oligomeric aggregate precursors^{45,46}. As these mice

were otherwise normal, these inclusions presumably comprise wild-type cellular proteins that are not aggregate-prone under conditions allowing autophagy, suggesting that these aggregates arose from the failure to clear soluble proteins, which aggregated when they exceeded critical levels. Thus, the reduction we have observed in both soluble species and aggregates in HD-cell and animal models treated with rapamycin could be explained by a model in which autophagy does not directly clear aggregates themselves but clears aggregate precursors. If these are removed, the equilibrium shifts away from aggregate formation. Even large aggregates can be degraded if synthesis of the mutant protein is switched off (for example, the switching off of transgene expression in an inducible mouse model⁵⁸ (Fig. 3)). Recent data have shown that huntingtin aggregates (which are much larger than typical autophagosomes) can be decorated with antibodies to the autophagosome-specific protein LC3 or with yellow-fluorescent-protein labelled LC3 (refs 18, 59). Although these studies did not test whether these inclusions were membrane-bounded, and this does not seem to be the case from most publications in the field⁶⁰, these observations raise the possibility that autophagic clearance of large aggregates may occur. In future, it will be important to demonstrate that these structures are indeed membrane-bounded and to quantify the contribution of this putative mechanism to the overall clearance of different aggregate-prone intracellular proteins.

In addition to rapamycin's protective effect, which is mediated by enhanced clearance of the primary toxins — the aggregate-prone proteins — the autophagy this drug induces may have other cytoprotective effects. Rapamycin pretreatment protects cells and *Drosophila* against various subsequent apoptotic insults⁴⁴ — a protective effect that was lost when autophagy was inhibited. However, this protection was confined to apoptotic models relying on mitochondrial-dependent cytochrome *c* release for the activation of downstream caspases, and was not seen in cells dying through the mitochondrial independent death-receptor-mediated pathway. Consistent with these data, no additional protective effect of rapamycin was observed when cytochrome *c* release from mitochondria was blocked. A plausible mechanism for this protective effect of rapamycin is that autophagy enhances mitochondrial clearance — we observed decreased levels of several mitochondrial proteins upon rapamycin pretreatment. After pro-apoptotic insults, we observed decreased levels of cytosolic cytochrome *c* and activated caspases in cells treated with rapamycin, consistent with reduced mitochondrial load. We proposed that the protective effect of rapamycin could be accounted for by enhanced clearance of mitochondria by autophagy, thereby reducing the release of cytosolic cytochrome *c* and activation of downstream caspases after pro-apoptotic insults⁴⁴. These data are also compatible with the suggestion that autophagy reduces susceptibility to oxidative stress⁶¹.

Pro-autophagic treatments might therefore have two distinct beneficial effects in protein conformational diseases. First, they could clear the primary toxin causing these diseases — the aggregate-prone protein. Second, enhanced autophagy attenuates apoptotic responses to various insults. In a number of neurodegenerative conditions, such insults may include excitotoxicity.

Upregulation of autophagy by huntingtin aggregates

The protective effects of autophagy upregulation that rapamycin treatment confers on cells containing protein aggregates seem to be mimicked by cells containing polyglutamine aggregates (in, for example, HD)²¹. In such cells, the aggregates sequester mTOR, reducing levels of the soluble protein. This results in decreased mTOR activity, and a consequent increase in autophagy. This phenomenon seems to be confined to cells with aggregates, and is not seen in cells expressing soluble mutant protein. The ability of cells to induce autophagy when they contain polyglutamine aggregates may account for the observations that such mutant-huntingtin-aggregate-containing cells are less prone to cell death than cells expressing non-aggregated mutant polyglutamine proteins⁶². This may be due to the more rapid clearance of the soluble protein in such cells, leading to lower levels of diffuse huntingtin (which seems to correlate with lower toxicity⁶²) and the protective effects of autophagy induction against subsequent pro-apoptotic insults⁴⁴.

No studies have shown polyglutamine disease aggregates to be as benign as the wild-type protein. Thus, even when aggregates seem to be less toxic than the soluble mutant protein⁶², they may be deleterious relative to the wild type. Regardless of which of these models is correct, reducing the load of toxic proteins in the cell is likely to be beneficial, particularly because the induction of autophagy reduces levels of both soluble and aggregated species.

Perspectives

The prospect that autophagy upregulation could be of therapeutic value to a range of neurodegenerative diseases caused by aggregate-prone proteins raises various challenges. Efficacy is likely to be greatest if treatment is started as early as possible, with the aim of delaying disease onset. In monogenic diseases such as HD (in which the phenotype is closely associated with a specific mutated gene), identifying at-risk individuals is feasible by screening for the mutation at the DNA level, as most cases will have family histories. If, in such cases, onset could be delayed until after normal life expectancy, disease would effectively have been prevented. It is also possible that degeneration might be more easily slowed if treatment is started earlier, providing another motivation for presymptomatic treatment approaches. However, the use of this strategy for the prevention of complex diseases would present the major challenge of identifying individuals at greatest risk unless a drug without serious side effects were used.

Rapamycin is designed for long-term human use (for example, to prevent renal transplant rejection⁶³) and rapamycin analogues are in phase III trials for the treatment of gliomas^{64,65}. Although rapamycin is the most specific kinase inhibitor known, its target, mTOR, controls many processes independent of autophagy⁶⁶. As a result, long-term rapamycin therapy is accompanied by side effects, including poor wound healing and a degree of immunosuppression. However, I am not aware of any side effects in humans that have been attributed to enhanced autophagy. So, although rapamycin may be worthy of consideration for the treatment of diseases such as HD, it would be desirable to find safer alternatives. These may include more specific targets downstream of mTOR (as it is not clear which of its substrates regulate mammalian autophagy), or the identification of mTOR-independent pathways. Recently, we found that autophagy can be regulated in an mTOR-independent manner by drugs that reduce intracellular inositol levels, such as lithium, valproate and carbamazepine⁶⁷. Our data suggest that greater levels of autophagy and aggregate-prone protein clearance can be induced by using drugs that target both pathways. Combination therapy with more moderate mTOR-independent and mTOR-dependent inhibition may be safer for long-term treatment than higher doses of drugs acting on either pathway. When used in combination, lower doses of each drug may achieve the same degree of overall autophagy induction, possibly providing a larger safety window.

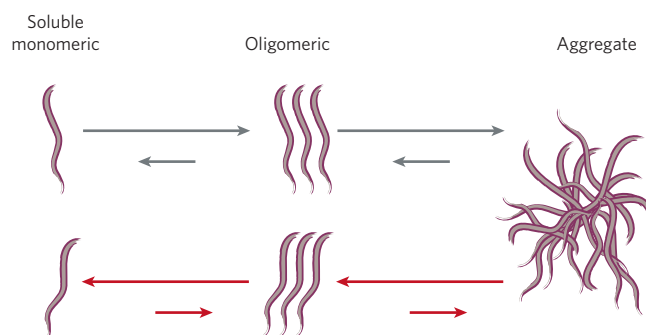


Figure 3 | Removal of aggregates can occur through removal of their precursors. Under normal conditions of substrate synthesis and basal autophagy, the direction of the equilibrium of an aggregate-prone protein is towards aggregate formation (grey arrows). If substrate synthesis is stopped (for example, in a conditional transgenic mouse) or clearance of the soluble/oligomeric forms is enhanced (for instance, by autophagy) then this equilibrium can be reversed and aggregates are indirectly cleared (red arrows).

It is difficult to be certain that prolonged upregulation of autophagy would be without risks. One theoretical concern is that enhanced mitochondrial turnover might be deleterious. Steady-state levels of mitochondria are set when the absolute degradation rate equals the absolute synthetic rate. If absolute synthetic rates are fixed, and autophagy is induced, this increases the fractional degradation rate (analogous to decreasing half-life). The cells will then adjust to a new steady state, in which the absolute degradation rate is maintained as it was previously. This can be explained because the mitochondrial pool decreases, and the absolute degradation rate is a function of concentration as well as fractional degradation rate. In other words, the absolute degradation rate can stay the same in the presence of enhanced autophagy (fractional degradation rate) when the mitochondrial load (concentration) drops. So, increasing autophagy decreases the mitochondrial load to a lower steady-state level but does not result in eventual total depletion of mitochondria (unless it is much more marked than the conditions we used)⁴⁴.

Very large decreases in mitochondrial load may be associated with deleterious effects due to a decrease in oxidative phosphorylation. However, it should be noted that although such relationships are complex and tissue-dependent, the activities of some respiratory complexes can be reduced by 25–80% before respiration or ATP synthesis in brain mitochondria are affected, and the activity of rat liver complex III can be decreased by 45% before respiration is affected⁶⁸. Thus, it is likely to be possible to induce autophagy and reduce mitochondrial load to levels that have substantial protective effects against proteinopathies but do not affect respiration.

Although rapamycin-induced autophagy is likely to be relatively safe in the long term, a degree of caution may be required when using drugs that act on other pathways, particularly if the drugs have not been designed or tested for chronic use. For instance, overexpression of ATG5 induces autophagy accompanied by cell death, the latter of which seems to be mediated by ATG5 interacting with Fas-associated protein with death domain (FADD)⁶⁹.

Finally, clinical trials of strategies that prevent neurodegeneration are generally costly, requiring hundreds of cases to be monitored over periods of years to generate sufficient data from which to detect moderate effects. If we want to be able to prioritize strategies and refine protocols for such trials, or if we want to undertake meaningful presymptomatic trials, then we need to define and characterize biomarkers that allow quantification of presymptomatic pathology and early disease progression in smaller patient cohorts over shorter time periods than are required for trials based on overt clinical signs. ■

- Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *Science* **296**, 1991–1995 (2002).
- Ross, C. A. & Poirier, M. A. What is the role of protein aggregation in neurodegeneration? *Nature Rev. Mol. Cell Biol.* **6**, 891–898 (2005).
- Perutz, M. F. & Windle, A. H. Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats. *Nature* **412**, 143–144 (2001).
- Pangalos, M. N., Jacobsen, S. J. & Reinhart, P. H. Disease modifying strategies for the treatment of Alzheimer's disease targeted at modulating levels of the β -amyloid peptide. *Biochem. Soc. Trans.* **33**, 553–558 (2005).
- Wellington, C. L. et al. Caspase cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease. *J. Neurosci.* **22**, 7862–7872 (2002).
- Luo, S., Vacher, C., Davies, J. E. & Rubinshtein, D. C. Cdk5 phosphorylation of huntingtin reduces its cleavage by caspases: implications for mutant huntingtin toxicity. *J. Cell Biol.* **169**, 647–656 (2005).
- Wellington, C. L. et al. Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J. Biol. Chem.* **273**, 9158–9167 (1998).
- Ciechanover, A. The ubiquitin proteolytic system: from a vague idea, through basic mechanisms, and onto human diseases and drug targeting. *Neurology* **66**, S7–S19 (2006).
- Richly, H. et al. A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* **120**, 73–84 (2005).
- Weihl, C. C., Dalal, S., Pestronk, A. & Hanson, P. I. Inclusion body myopathy-associated mutations in p97/VCP impair endoplasmic reticulum-associated degradation. *Hum. Mol. Genet.* **15**, 189–199 (2006).
- Ravikumar, B., Duden, R. & Rubinshtein, D. C. Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. *Hum. Mol. Genet.* **11**, 1107–1117 (2002).
- Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N. & Rubinshtein, D. C. α -Synuclein is degraded by both autophagy and the proteasome. *J. Biol. Chem.* **278**, 25009–25013 (2003).
- Verhoef, L. G., Lindsten, K., Masucci, M. G. & Dantuma, N. P. Aggregate formation inhibits proteasomal degradation of polyglutamine proteins. *Hum. Mol. Genet.* **11**, 2689–2700 (2002).
- Venkataraman, P., Wetzel, R., Tanaka, M., Nukina, N. & Goldberg, A. L. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol. Cell* **14**, 95–104 (2004).
- Holmberg, C. I., Stanislawski, K. E., Mensah, K. N., Matouschek, A. & Morimoto, R. I. Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J.* **23**, 4307–4318 (2004).
- Yorimitsu, T. & Klionsky, D. J. Autophagy: molecular machinery for self-eating. *Cell Death Differ.* **12**, 1542–1552 (2005).
- Shibata, M. et al. Regulation of intracellular accumulation of mutant Huntingtin by Beclin 1. *J. Biol. Chem.* **281**, 14474–14485 (2006).
- Iwata, A. et al. Increased susceptibility of cytoplasmic over nuclear polyglutamine aggregates to autophagic degradation. *Proc. Natl Acad. Sci. USA* **102**, 13135–13140 (2005).
- Qin, Z. H. et al. Autophagy regulates the processing of amino terminal huntingtin fragments. *Hum. Mol. Genet.* **12**, 3231–3244 (2003).
- Berger, Z. et al. Rapamycin alleviates toxicity of different aggregate-prone proteins. *Hum. Mol. Genet.* **15**, 433–442 (2006).
- Ravikumar, B. et al. mTOR inhibition induces autophagy and reduces toxicity of the Huntington's disease mutation in *Drosophila* and mouse models. *Nature Genet.* **36**, 585–595 (2004).
- Komatsu, M. et al. Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Ravikumar, B. et al. Raised intracellular glucose concentrations reduce aggregation and cell death caused by mutant huntingtin exon 1 by decreasing mTOR phosphorylation and inducing autophagy. *Hum. Mol. Genet.* **12**, 985–994 (2003).
- Rideout, H., Lang-Rollin, I. & Stefanis, L. Involvement of macroautophagy in the dissolution of neuronal inclusions. *Int. J. Biochem. Cell Biol.* **36**, 2551–2562 (2004).
- Massey, A., Kiffin, R. & Cuervo, A. M. Pathophysiology of chaperone-mediated autophagy. *Int. J. Biochem. Cell Biol.* **36**, 2420–2434 (2004).
- McNaught, K. S., Belzair, R., Isacson, O., Jenner, P. & Olanow, C. W. Altered proteasomal function in sporadic Parkinson's disease. *Exp. Neurol.* **179**, 38–46 (2003).
- Seo, H., Sonntag, K. C. & Isacson, O. Generalized brain and skin proteasome inhibition in Huntington's disease. *Ann. Neurol.* **56**, 319–328 (2004).
- Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
- Bowman, A. B., Yoo, S. Y., Dantuma, N. P. & Zoghbi, H. Y. Neuronal dysfunction in a polyglutamine disease model occurs in the absence of ubiquitin-proteasome system impairment and inversely correlates with a degree of nuclear inclusion formation. *Hum. Mol. Genet.* **14**, 679–691 (2005).
- Diaz-Hernandez, M. et al. Neuronal induction of the immunoproteasome in Huntington's disease. *J. Neurosci.* **23**, 11653–11661 (2003).
- Sun, X. M. et al. Caspase activation inhibits proteasome function during apoptosis. *Mol. Cell* **14**, 81–93 (2004).
- Zhang, Y. et al. Parkin functions as an E2-dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc. Natl Acad. Sci. USA* **97**, 13354–13359 (2000).
- Shimura, H. et al. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nature Genet.* **25**, 302–305 (2000).
- Imai, Y., Soda, M. & Takahashi, R. Parkin suppresses unfolded protein stress-induced cell death through its E3 ubiquitin-protein ligase activity. *J. Biol. Chem.* **275**, 35661–35664 (2000).
- Leroy, E. et al. The ubiquitin pathway in Parkinson's disease. *Nature* **395**, 451–452 (1998).
- Moore, D. J., West, A. B., Dawson, V. L. & Dawson, T. M. Molecular pathophysiology of Parkinson's disease. *Annu. Rev. Neurosci.* **28**, 57–87 (2005).
- Healy, D. G. et al. UCHL-1 is not a Parkinson's disease susceptibility gene. *Ann. Neurol.* **59**, 627–633 (2006).
- Liu, Y., Fallon, L., Lashuel, H. A., Liu, Z. & Lansbury, P. T. The UCH-L1 gene encodes two opposing enzymatic activities that affect α -synuclein degradation and Parkinson's disease susceptibility. *Cell* **111**, 209–218 (2002).
- Saigoh, K. et al. Intragenic deletion in the gene encoding ubiquitin carboxy-terminal hydrolase in *gad* mice. *Nature Genet.* **23**, 47–51 (1999).
- Watts, G. D. et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature Genet.* **36**, 377–381 (2004).
- McNaught, K. S., Perl, D. P., Brownell, A. L. & Olanow, C. W. Systemic exposure to proteasome inhibitors causes a progressive model of Parkinson's disease. *Ann. Neurol.* **56**, 149–162 (2004).
- Kuma, A. et al. The role of autophagy during the early neonatal starvation period. *Nature* **432**, 1032–1036 (2004).
- Boya, P. et al. Inhibition of macroautophagy triggers apoptosis. *Mol. Cell Biol.* **25**, 1025–1040 (2005).
- Ravikumar, B., Berger, Z., Vacher, C., O'Kane, C. J. & Rubinshtein, D. C. Rapamycin pre-treatment protects against apoptosis. *Hum. Mol. Genet.* **15**, 1209–1216 (2006).
- Komatsu, M. et al. Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441**, 880–884 (2006).
- Hara, T. et al. Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
- Ravikumar, B. et al. Dynein mutations impair autophagic clearance of aggregate-prone proteins. *Nature Genet.* **37**, 771–776 (2005).
- Levy, J. R. et al. A motor neuron disease-associated mutation in p150^{Glued} perturbs dynein function and induces protein aggregation. *J. Cell Biol.* **172**, 733–745 (2006).
- Puls, I. et al. Mutant dynein in motor neuron disease. *Nature Genet.* **33**, 455–456 (2003).
- Hafezparast, M. et al. Mutations in dynein link motor neuron degeneration to defects in retrograde transport. *Science* **300**, 808–812 (2003).
- Kieran, D. et al. A mutation in dynein rescues axonal transport defects and extends the life span of ALS mice. *J. Cell Biol.* **169**, 561–567 (2005).

52. Ligon, L. A. *et al.* Mutant superoxide dismutase disrupts cytoplasmic dynein in motor neurons. *Neuroreport* **16**, 533–536 (2005).
53. Cuervo, A. M., Stefanski, L., Fredenburg, R., Lansbury, P. T. & Sulzer, D. Impaired degradation of mutant α -synuclein by chaperone-mediated autophagy. *Science* **305**, 1292–1295 (2004).
54. Harper, S. Q. *et al.* RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc. Natl Acad. Sci. USA* **102**, 5820–5825 (2005).
55. Rodriguez-Lebron, E. & Paulson, H. L. Allele-specific RNA interference for neurological disease. *Gene Ther.* **13**, 576–581 (2006).
56. Davies, J. E., Sarkar, S. & Rubinshtein, D. C. Trehalose reduces aggregate formation and delays pathology in a transgenic mouse model of oculopharyngeal muscular dystrophy. *Hum. Mol. Genet.* **15**, 23–31 (2006).
57. Sanchez, I., Mahlke, C. & Yuan, Y. Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. *Nature* **421**, 373–379 (2003).
58. Martin-Aparicio, E. *et al.* Proteasomal-dependent aggregate reversal and absence of cell death in a conditional mouse model of Huntington's disease. *J. Neurosci.* **21**, 8772–8781 (2001).
59. Yamamoto, A., Cremona, M. L. & Rothman, J. E. Autophagy-mediated clearance of huntingtin aggregates triggered by the insulin-signaling pathway. *J. Cell Biol.* **172**, 719–731 (2006).
60. DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993 (1997).
61. Kiffin, R., Bandyopadhyay, U. & Cuervo, A. Oxidative stress and autophagy. *Antioxid. Redox Signal.* **8**, 152–162 (2006).
62. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805–810 (2004).
63. Lee, V. W. & Chapman, J. R. Sirolimus: its role in nephrology. *Nephrology (Carlton)* **10**, 606–614 (2005).
64. Bjornsti, M. A. & Houghton, P. J. The TOR pathway: a target for cancer therapy. *Nature Rev. Cancer* **4**, 335–348 (2004).
65. Galanis, E. *et al.* Phase II trial of temsirolimus (CCI-779) in recurrent glioblastoma multiforme: a North Central Cancer Treatment Group Study. *J. Clin. Oncol.* **23**, 5294–5304 (2005).
66. Wullschlegel, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124**, 471–484 (2006).
67. Sarkar, S. *et al.* Lithium induces autophagy by inhibiting inositol monophosphatase. *J. Cell Biol.* **170**, 1101–1111 (2005).
68. Murphy, M. P. How understanding the control of energy metabolism can help investigation of mitochondrial dysfunction, regulation and pharmacology. *Biochim. Biophys. Acta* **1504**, 1–11 (2001).
69. Pyo, J. *et al.* Essential roles of Atg5 and FADD in autophagic cell death: dissection of autophagic cell death into vacuole formation and cell death. *J. Biol. Chem.* **280**, 20722–20728 (2005).
70. Nasir, J. *et al.* Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioural and morphological changes in heterozygotes. *Cell* **81**, 811–823 (1995).
71. Zeitlin, S. *et al.* Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homolog. *Nature Genet.* **11**, 155–163 (1995).
72. Duyao, M. P. *et al.* Inactivation of the mouse Huntington's disease gene homolog *Hdh*. *Science* **269**, 407–410 (1995).
73. Harper, P. S. *Huntington's Disease* 2nd edn (WB Saunders, London, 1996).
74. Rubinshtein, D. C. Lessons from animal models of Huntington's disease. *Trends Genet.* **18**, 202–209 (2002).
75. Ordway, J. M. *et al.* Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* **91**, 753–763 (1997).
76. Cattaneo, E., Zuccato, C. & Tartari, M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nature Rev. Neurosci.* **6**, 919–930 (2005).

Acknowledgements I am grateful to B. Ravikumar and M. Futter for critical comments on the manuscript and L. Smith for help with manuscript preparation. The work in my laboratory covered by this review has been funded by a Wellcome Trust Senior Fellowship in Clinical Science, a Medical Research Council (MRC) Programme Grant, Wyeth, and European Union Framework VI (EUROSCA).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to the author (dcrl000@hermes.cam.ac.uk).

Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases

Michael T. Lin¹ & M. Flint Beal¹

Many lines of evidence suggest that mitochondria have a central role in ageing-related neurodegenerative diseases. Mitochondria are critical regulators of cell death, a key feature of neurodegeneration. Mutations in mitochondrial DNA and oxidative stress both contribute to ageing, which is the greatest risk factor for neurodegenerative diseases. In all major examples of these diseases there is strong evidence that mitochondrial dysfunction occurs early and acts causally in disease pathogenesis. Moreover, an impressive number of disease-specific proteins interact with mitochondria. Thus, therapies targeting basic mitochondrial processes, such as energy metabolism or free-radical generation, or specific interactions of disease-related proteins with mitochondria, hold great promise.

Neurodegenerative diseases are a heterogeneous group of disorders characterized by gradually progressive, selective loss of anatomically or physiologically related neuronal systems. Prototypical examples include Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and Huntington's disease (HD). Despite this heterogeneity, we argue that mitochondrial involvement is likely to be an important common theme in these diseases. Mitochondria are key regulators of cell survival and death (Fig. 1; for a review see ref. 1), have a central role in ageing, and have recently been found to interact with many of the specific proteins implicated in genetic forms of neurodegenerative diseases (Table 1).

Mitochondria and ageing

By far the greatest risk factor for neurodegenerative diseases such as AD, PD and ALS is ageing, and mitochondria have been thought to contribute to ageing through the accumulation of mitochondrial DNA (mtDNA) mutations and net production of reactive oxygen species (ROS).

Although most mitochondrial proteins are encoded by the nuclear genome, mitochondria contain many copies of their own DNA. Human mtDNA is a circular molecule of 16,569 base pairs that encodes 13 polypeptide components of the respiratory chain, as well as the rRNAs and tRNAs necessary to support intramitochondrial protein synthesis

using its own genetic code. Inherited mutations in mtDNA are known to cause a variety of diseases, most of which affect the brain and muscles — tissues with high energy requirements. One hypothesis has been that somatic mtDNA mutations acquired during ageing contribute to the physiological decline that occurs with ageing and ageing-related neurodegeneration.

It is well established that mtDNA accumulates mutations with ageing, especially large-scale deletions² and point mutations. In the mtDNA control region, point mutations at specific sites can accumulate to high levels in certain tissues: T414G in cultured fibroblasts, A189G and T408A in muscle, and C150T in white blood cells³. However, these control-region 'hot spots' have not been observed in the brain⁴. Point mutations at individual nucleotides seem to occur at low levels in the brain⁵, although the overall level may be high. Using a polymerase chain reaction (PCR)-cloning-sequencing strategy, we found that the average level of point mutations in two protein-coding regions of brain mtDNA from elderly subjects was ~2 mutations per 10 kb⁶. Noncoding regions, which may be under less selection pressure, potentially accumulate between twice and four times as many⁷.

The accumulation of these deletions and point mutations with ageing correlates with decline in mitochondrial function. For example, a negative correlation has been found between brain cytochrome oxidase

Table 1 | Proteins that have a function in major neurodegenerative diseases with mitochondrial involvement

Disease	Genetic causes	Function
Alzheimer's disease	APP	Gives rise to A β , the primary component of senile plaques
	PS1 and PS2	A component of γ -secretase, which cleaves APP to yield A β
Parkinson's disease	α -Synuclein	The primary component of Lewy bodies
	Parkin	A ubiquitin E3 ligase
	DJ-1	Protects the cell against oxidant-induced cell death
	PINK1	A kinase localized to mitochondria. Function unknown. Seems to protect against cell death
	LRRK2	A kinase. Function unknown
Amyotrophic lateral sclerosis	HTRA2	A serine protease in the mitochondrial intermembrane space. Degrades denatured proteins within mitochondria. Degrades inhibitor of apoptosis proteins and promotes apoptosis if released into the cytosol
	SOD1	Converts superoxide to hydrogen peroxide. Disease-causing mutations seem to confer a toxic gain of function
Huntington's disease	Huntingtin	Function unknown. Disease-associated mutations produce expanded polyglutamine repeats

¹Department of Neurology and Neuroscience, Weill Medical College of Cornell University, Room F-610, 525 East 68th Street, New York 10021, USA.

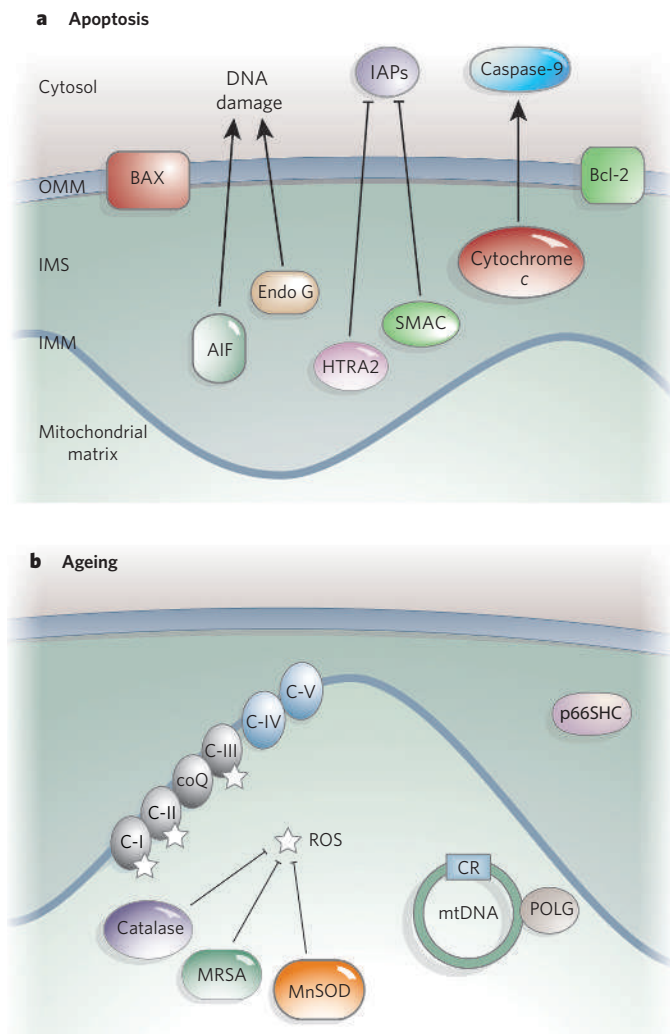


Figure 1 | Role of mitochondria in apoptosis and ageing. **a**, Several intermembrane space (IMS) proteins are pro-apoptotic if released into the cytosol. Cytochrome *c* (C) activates caspase-9. SMAC (second mitochondrial activator of caspases) and HTRA2 inhibit cytosolic inhibitors of apoptosis proteins (IAPs). HTRA2 is a serine protease that might function to remove denatured proteins within mitochondria, but degrades IAPs when released from mitochondria. Inhibiting HTRA2's normal quality control function or enhancing its IAP-degrading activity could both promote cell death. Apoptosis inhibitor factor (AIF) and endonuclease G (endo G) translocate to the nucleus and induce chromatin condensation and DNA fragmentation. Release of these proteins into the cell is modulated by recruitment of BAX (which is proapoptotic) or Bcl-2 (anti-apoptotic) to the outer mitochondrial membrane (OMM). Numerous extracellular and intracellular signals converge to regulate mitochondrial apoptosis (reviewed in ref. 1; see also page 796). **b**, Mitochondrial DNA (mtDNA) mutations and oxidative stress may contribute to ageing. Somatic mutations accumulate in mtDNA with age. Inducing mtDNA mutations by disabling the proofreading activity of mtDNA POLG accelerates ageing-related pathology in transgenic mice. Certain mtDNA polymorphisms are associated with increased longevity, possibly by reducing membrane potential and decreasing the generation of reactive oxygen species (ROS, white stars). Overexpression of ROS-scavenging enzymes manganese superoxide dismutase (MnSOD), methionine sulphoxide reductase A (MSRA) or catalase within mitochondria prolongs lifespan. Knockout of p66SHC, a protein that promotes ROS generation and mitochondrial apoptosis, also prolongs lifespan. Complex IV (C-IV) and complex V activities decline with ageing, and knockdown of complex V activity causes oxidative damage to nuclear DNA, which probably results in decreased gene expression with ageing. IMM, inner mitochondrial membrane.

activity and increased point-mutation levels in a cytochrome oxidase gene (*COI*)⁶. Moreover, somatic deletions can be clonally expanded in individual neurons, and high levels of such deletions correlate with cytochrome oxidase deficiency on a cell-by-cell basis in the substantia nigra, perhaps contributing to the age dependence of PD^{8,9}. However, although the cell-by-cell correlation provides strong circumstantial evidence, correlations do not prove that somatic mtDNA mutations cause age-related pathology.

Recently, several groups have addressed the issue of causation using a clever approach to generate mtDNA mutations experimentally (for a review see ref. 10). MtDNA replication is carried out by mtDNA polymerase- γ (POLG), which has 3'-to-5' exonuclease (proofreading) activity in addition to its 5'-to-3' polymerase activity. If the proofreading activity of POLG is eliminated and the polymerase activity preserved, mtDNA mutations accumulate because of uncorrected errors during replication. In mice with such proofreading-deficient POLG (mtDNA-mutator mice), mtDNA mutations accumulate to high levels in all tissues. By 8 weeks of age, homozygous *Polg*^{-/-} animals had 9 point mutations per 10 kb in cytochrome *b*. By contrast, normal mice had less than 1 mutation per 10 kb. This marked increase in mtDNA mutations resulted in decreased respiratory enzyme activity and ATP production. To begin with, the mice appeared normal, but by 25 weeks of age began to exhibit pathology frequently seen in human (although not necessarily murine) ageing, including weight loss, alopecia, osteoporosis, kyphosis, cardiomyopathy, anaemia, gonadal atrophy and sarcopaenia. The median lifespan of such mice was 48 weeks (none lived beyond 61 weeks of age) — much shorter than the typical murine lifespan of 2 years.

A second, independent mtDNA-mutator mouse showed a similar marked increase in mtDNA mutations, progeric features and early mortality. Notably, neuropathology was not reported in either mouse model, although detailed examination was not carried out. Humans with *POLG* mutations exhibit parkinsonism, ophthalmoplegia and myopathy (see below), and the reasons for the differences between mice and humans with such mutations are not yet known.

In both mtDNA-mutator mouse models, markers of apoptosis such as activated caspase 3 were increased at times coinciding with tissue degeneration, suggesting that apoptosis mediates deleterious effects of somatic mtDNA mutations. Interestingly, tissues from the mtDNA-mutator mice did not show increased levels of lipid, protein or DNA oxidation, hydrogen peroxide production, or sensitivity to oxidative stress. Thus, the effects of mtDNA mutations in these mice do not seem to be mediated through ROS production.

Net production of ROS is another important mechanism by which mitochondria are thought to contribute to ageing. Mitochondria contain multiple electron carriers capable of producing ROS, as well as an extensive network of antioxidant defences (Fig. 2). Mitochondrial insults, including oxidative damage itself, can cause an imbalance between ROS production and removal, resulting in net ROS production¹¹. The importance to ageing of net mitochondrial ROS production is supported by observations that enhancing mitochondrial antioxidant defences can increase longevity. In *Drosophila*, overexpression of the mitochondrial antioxidant enzymes manganese superoxide dismutase (MnSOD)¹² and methionine sulphoxide reductase¹³ prolongs lifespan. This strategy is most successful in short-lived strains of *Drosophila*, and has no effect in already long-lived strains. However, it has recently been shown that overexpression of catalase experimentally targeted to mitochondria increased lifespan in an already long-lived mouse strain¹⁴. The authors of this work generated transgenic mice overexpressing catalase targeted to peroxisomes, nuclei or mitochondria. The mitochondrially targeted construct provided the maximal benefit, increasing median and maximal lifespan by 20%. Hydrogen peroxide production and oxidative inactivation of aconitase were reduced in isolated cardiac mitochondria, DNA oxidation and levels of mitochondrial deletions were reduced in skeletal muscle, and cardiac pathology, arteriosclerosis and cataract development were delayed.

In humans, a recent study of gene expression in the brain suggests that oxidative damage has a major role in the cognitive decline that

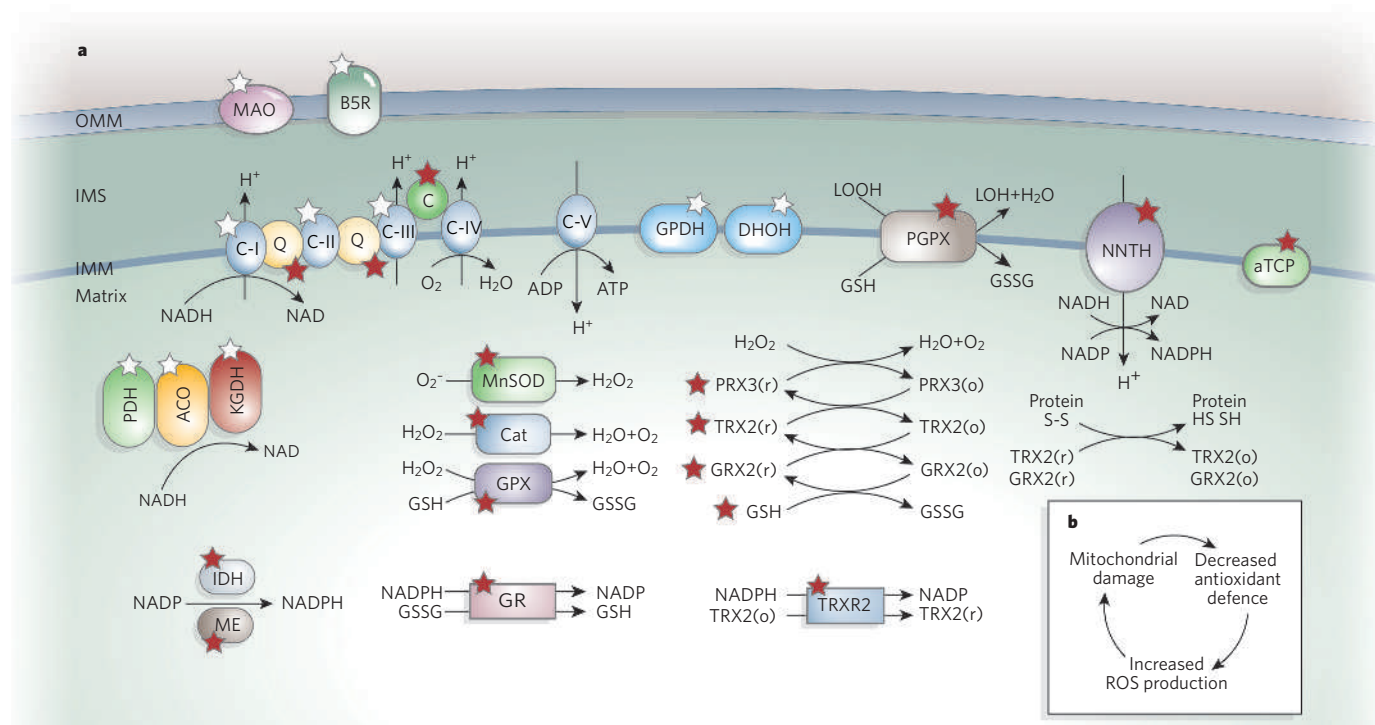


Figure 2 | Role of mitochondria in reactive oxygen species metabolism. The processes and components involved in ROS generation (white stars) and antioxidant defence (red stars). Mitochondria are the primary cellular consumers of oxygen and contain numerous redox enzymes capable of transferring single electrons to oxygen, generating the ROS superoxide ($O_2^{\cdot -}$). Mitochondrial enzymes so far known to generate ROS include the tricarboxylic acid (TCA) cycle enzymes aconitase (ACO) and α -ketoglutarate dehydrogenase (KGDH); the electron-transport chain (ETC) complexes I, II and III; pyruvate dehydrogenase (PDH) and glycerol-3-phosphate dehydrogenase (GPDH); dihydroorotate dehydrogenase (DHOH); the monoamine oxidases (MAO) A and B; and cytochrome b_5 reductase (B5R). The transfer of electrons to oxygen, generating superoxide, is more likely when these redox carriers are abundantly charged with electrons and the potential energy for transfer is high, as reflected by a high mitochondrial membrane potential. ROS generation is decreased when available electrons are few and potential energy for the transfer is low. Mitochondria also contain an extensive antioxidant defence system to detoxify the ROS generated by the reactions described above. Both the membrane-enclosed and soluble compartments are protected. Nonenzymatic components of the system include α -tocopherol (aTCP), coenzyme Q10 (Q), cytochrome c (C) and glutathione (GSH). Enzymatic components include manganese superoxide dismutase (MnSOD), catalase (Cat), glutathione peroxidase (GPX), phospholipid hydroperoxide glutathione peroxidase (PGPX), glutathione reductase (GR); peroxiredoxins (PRX3/5), glutaredoxin (GRX2), thioredoxin (TRX2) and thioredoxin reductase (TRXR2). The regeneration of GSH (through GR) and reduced TRX2 (through TRXR2) depends on NADPH, which is derived from substrates (through isocitrate dehydrogenase, IDH, or malic enzyme, ME) or the membrane potential (through nicotinamide nucleotide transhydrogenase, NNTH). So, like ROS generation, antioxidant defences are also tied to the redox and energetic state of mitochondria. GSSG, glutathione disulphide; LOH, lipid hydroxide; LOOH, lipid hydroperoxide; o, oxidized state; r, reduced state. In structurally and functionally intact mitochondria, a large antioxidant defence capacity balances ROS generation, and there is little net ROS production. Mitochondrial damage with decrease of antioxidant defence capacity is a prerequisite for net ROS production. Once this occurs, a vicious cycle (inset) can ensue whereby ROS can further damage mitochondria, causing more free-radical generation and loss or consumption of antioxidant capacity. For example, the Fe-S cluster in aconitase is easily inactivated by superoxide, the iron is released, and this induces hydroxyl radical production. (For a review of the role of mitochondria in ROS metabolism, see ref. 11.)

accompanies ageing¹⁵. Transcriptional profiling of postmortem frontal cortex samples from individuals aged from 26 to 106 revealed that after the age of 40 there was a decrease in the expression of genes involved in synaptic plasticity, vesicular transport and mitochondrial function, followed by increased expression of stress-response, antioxidant and DNA-repair genes. In the brain, the age-downregulated genes suffered markedly increased oxidative DNA damage compared with the age-stable or age-upregulated genes. Promoter regions were particularly affected, perhaps because they contain G/C-rich sequences that are sensitive to oxidation, or do not undergo transcription-coupled repair. In SH-SY5Y cells, promoters of the same age-downregulated genes were both more sensitive to hydrogen-peroxide-induced damage and less able to undergo base excision repair of such damage than promoters of age-stable or age-upregulated genes.

To investigate whether impaired mitochondrial function could predispose these age-downregulated genes to DNA damage, small interfering RNA (siRNA) was used to reduce the expression of mitochondrial F_1 -ATPase 2.5-fold in SH-SY5Y cells, approximating the reduction seen in the aged human cortex. This resulted in significantly increased promoter DNA damage in age-downregulated genes, which was partly

reversed by the antioxidant vitamin E. These findings support the idea that mitochondrial dysfunction contributes to the damage of vulnerable genes in the ageing brain. The vulnerable-gene promoters are both more sensitive to oxidative stress and deficient in repair, and mitochondrial dysfunction could potentially exacerbate both by increasing ROS or decreasing the availability of ATP, which is necessary for repair.

Mitochondria and Alzheimer's disease

AD is characterized clinically by progressive cognitive decline, and pathologically by the presence of senile plaques composed primarily of amyloid- β peptide ($A\beta$) and neurofibrillary tangles made up mainly of hyperphosphorylated tau. About 5–10% of cases are familial, occurring in an early-onset, autosomal-dominant manner. Three proteins are known to be associated with such familial cases: amyloid precursor protein (APP) — which is cleaved sequentially by β - and γ -secretases to produce $A\beta$ — and presenilins 1 and 2 (PS1 and PS2), one or other of which is a component of each γ -secretase complex.

There is extensive literature supporting a role for mitochondrial dysfunction and oxidative damage in the pathogenesis of AD. Oxidative damage occurs early in the AD brain, before the onset of significant

plaque pathology¹⁶. Oxidative damage also precedes A β deposition in transgenic APP mice¹⁷, with upregulation of genes relating to mitochondrial metabolism and apoptosis occurring even earlier and co-localizing with the neurons undergoing oxidative damage¹⁸.

Moreover, such oxidative damage and mitochondrial dysfunction probably contribute causally to AD-related pathology. In fetal guinea pig neurons, hydrogen peroxide treatment increased intracellular A β levels¹⁹. Treatment with the mitochondrial uncoupler CCCP (carbonyl cyanide m-chlorophenylhydrazone) caused cultured astrocytes to mimic the amyloidogenic APP processing and intracellular A β accumulation that is seen in Down syndrome astrocytes²⁰. In a transgenic APP-mutant mouse, hemizygous deficiency of the mitochondrial antioxidant enzyme MnSOD markedly increased brain A β levels and plaque deposition²¹. In another transgenic APP-mutant mouse, energy metabolism inhibitors (insulin, 2-deoxyglucose, 3-nitropropionic acid and kainic acid) elevated β -secretase levels and activity and A β levels²².

Several pathways connecting oxidative stress and AD pathology have recently been uncovered. Oxidative stress may activate signalling pathways that alter APP or tau processing. For example, oxidative stress increases the expression of β -secretase through activation of c-Jun amino-terminal kinase and p38 mitogen-activated protein kinase (MAPK)²³, and increases aberrant tau phosphorylation by activation of glycogen synthase kinase 3 (ref. 24). Oxidant-induced inactivation of critical molecules may also be important. In a proteomic study, the prolyl isomerase PIN1 was found to be particularly sensitive to oxidative damage²⁵. PIN1 catalyses protein conformational changes that affect both APP and tau processing. Knockout of *Pin1* increases amyloidogenic APP processing and intracellular A β levels in mice²⁶. *Pin1*-knockout mice also exhibit tau hyperphosphorylation, motor and behavioural deficits, and neuronal degeneration²⁷.

There is some evidence that mtDNA may be involved in the mitochondrial dysfunction seen in AD. When patient mtDNA is transferred into mtDNA-deficient cell lines, the resulting 'cybrids' reproduce the respiratory enzyme deficiency seen in the brain and other tissues in AD, suggesting that the defect is carried at least in part by mtDNA abnormalities²⁸. However, identifying AD-specific mtDNA mutations has been a challenge. Complete sequencing of mtDNA from 145 AD patients and 128 controls did not reveal any significant association with mitochondrial haplogroup or with inherited mtDNA mutations²⁹. There was also no association with acquired mtDNA mutations when a coding region (for CO1) was examined⁶. However, in the same way that promoters appeared more sensitive to damage than coding regions in nuclear genes¹⁵, the mtDNA control region showed an increase in acquired mutations in AD³⁰. AD brains had on average a 63% increase in heteroplasmic mtDNA control-region mutations, and in individuals older than 80 years there was a 130% increase in mutations. These mutations preferentially altered known mtDNA regulatory elements and suppressed mitochondrial transcription and replication.

Finally, several recent reports suggest that many of the proteins implicated in AD pathogenesis have direct physical involvement with mitochondria or mitochondrial proteins (Fig. 3). APP has been found to have a dual endoplasmic reticulum/mitochondrial-targeting sequence, and in transfected cells and transgenic mice overexpressing APP it clogged the mitochondrial protein importation machinery, causing mitochondrial dysfunction and impaired energy metabolism³¹. A β binds to a mitochondrial-matrix protein termed A β -binding alcohol dehydrogenase (ABAD)³². Blocking the interaction of A β and ABAD with a 'decoy peptide' suppressed A β -induced apoptosis and free-radical generation in neurons. Conversely, overexpression of ABAD in transgenic APP-mutant mice exaggerated neuronal oxidative stress and impaired memory. Two other groups have also reported that A β interacts with mitochondria, inhibiting cytochrome oxidase activity and increasing free-radical generation^{33,34}. A β also inhibits α -ketoglutarate dehydrogenase activity in isolated mitochondria³⁵, and deficiency of α -ketoglutarate dehydrogenase³⁶ and cytochrome oxidase activities³⁷ has previously been observed in the brain and other tissues in AD. A β also interacts with the serine protease HTRA2 (also known as OMI)³⁸.

Presenilin and all the other components of the γ -secretase complex have also been localized to mitochondria, where they form an active γ -secretase complex³⁹.

Mitochondria and Parkinson's disease

PD is characterized clinically by progressive rigidity, bradykinesia and tremor, and pathologically by loss of pigmented neurons in the substantia nigra and the presence of Lewy bodies — distinctive cytoplasmic inclusions that immunostain for α -synuclein and ubiquitin.

Mitochondria were first implicated in PD because MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), whose metabolite MPP⁺ inhibits complex I of the mitochondrial electron-transport chain, caused parkinsonism in designer-drug abusers. This model has since been refined in laboratory animals, in which chronic infusion of rotenone⁴⁰ — another complex-I inhibitor — or MPTP⁴¹ results clinically in a parkinsonian phenotype and pathologically in nigral degeneration with cytoplasmic inclusions immunoreactive for α -synuclein and ubiquitin. The mechanism of toxicity in these complex-I inhibition models probably involves oxidative stress⁴². Complex-I inhibition and oxidative stress were shown to be relevant to naturally occurring PD when complex-I deficiency and glutathione depletion were found in the substantia nigra of patients with idiopathic PD and in patients with pre-symptomatic PD⁴³.

Many of the genes associated with PD also implicate mitochondria in disease pathogenesis. So far, mutations or polymorphisms in mtDNA and at least nine named nuclear genes have been identified as causing PD or affecting PD risk: α -synuclein, parkin, ubiquitin carboxy-terminal hydrolase L1, *DJ-1*, phosphatase and tensin homologue (PTEN)-induced kinase 1 (*PINK1*), leucine-rich-repeat kinase 2 (*LRRK2*), the nuclear receptor *NURR1*, *HTRA2* and tau. Of the nuclear genes, α -synuclein, parkin, *DJ-1*, *PINK1*, *LRRK2* and *HTRA2* directly or indirectly involve mitochondria.

In a small number of cases, inherited mtDNA mutations result in parkinsonism, typically as one feature of a larger syndrome. In one family, we found that the Leber's optic atrophy G11778A mutation was associated with L-DOPA-responsive parkinsonism, variably co-occurring with dementia, dystonia, ophthalmoplegia and ataxia⁴⁴. Notably, this mutation is in a subunit of complex I. Mutations in the nuclear-encoded mtDNA polymerase- γ (*POLG*) gene impair mtDNA replication and result in multiple mtDNA deletions, typically causing chronic progressive external ophthalmoplegia and myopathy. In such families, *POLG* mutations also cosegregate with parkinsonism⁴⁵.

There is less evidence for mtDNA involvement in non-syndromic PD. Nigral neurons from PD patients contain increased levels of clonally expanded somatic mtDNA deletions compared with those from age-matched controls, although high levels are also seen in normal ageing^{8,9}. We found no difference between PD and control subjects in inherited or acquired complex-I or tRNA point mutations^{46,47}. Interestingly, however, several groups have found that certain continent-specific clusters of polymorphisms, termed mtDNA haplogroups, may decrease the risk of developing PD. Among Europeans, the haplogroup cluster UJKT is associated with a decreased risk for PD compared with haplogroup H (ref. 48). It is of note that haplogroups underrepresented in PD patients are overrepresented in healthy centenarians⁴⁹. Protective mtDNA lineages seem to have arisen from areas requiring cold-adaptation, including relative uncoupling of mitochondria to increase heat generation at the expense of ATP production. It has been proposed that this partial uncoupling increases longevity and decreases risk of neurodegeneration by decreasing free-radical generation⁵⁰.

Mutations in α -synuclein are associated with autosomal dominant familial PD. α -Synuclein is a major component of Lewy bodies, and the primary effect of α -synuclein mutations is likely to be an increased formation of oligomeric or fibrillar aggregates. However, there seem to be close interrelationships between abnormal protein accumulation or degradation, oxidative stress and mitochondrial dysfunction. In transgenic mice, overexpression of α -synuclein impairs mitochondrial function, increases oxidative stress and enhances nigral pathology induced by MPTP⁵¹. Moreover, in a recent study of mice overexpressing

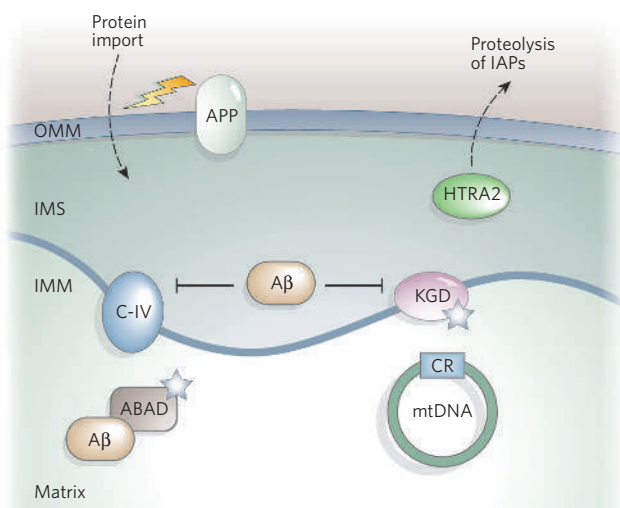
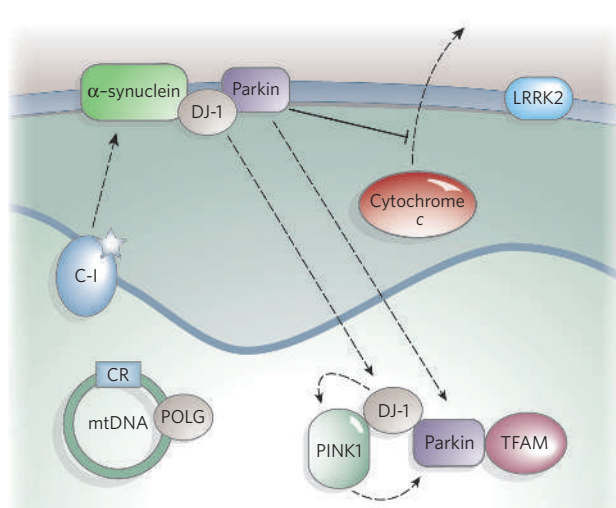
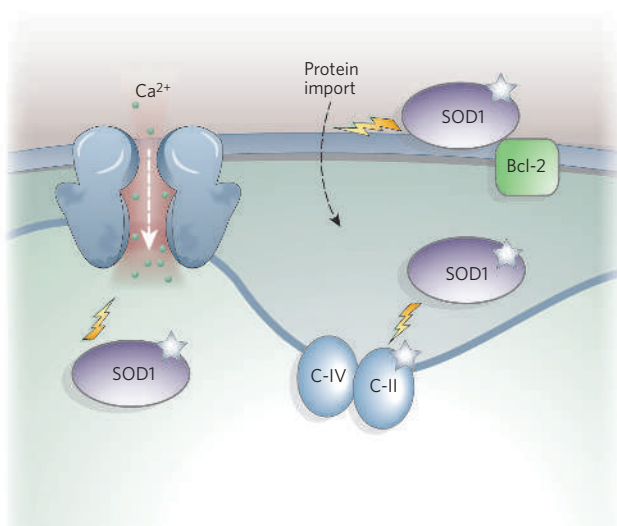
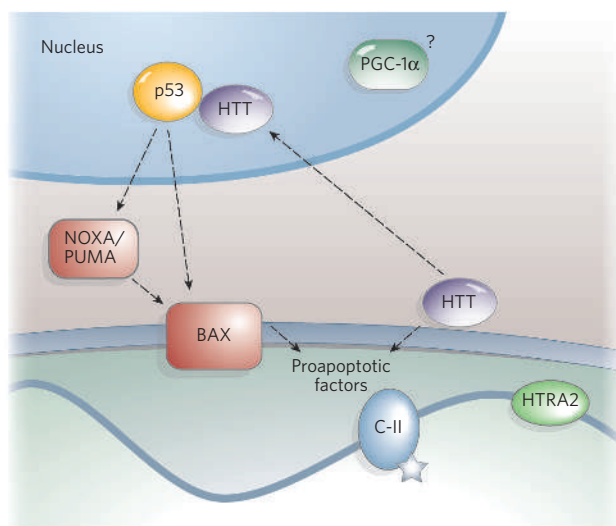
a Alzheimer's disease**b Parkinson's disease****c Amyotrophic lateral sclerosis****d Huntington's disease**

Figure 3 | The role of mitochondria in ageing-related neurodegenerative diseases. **a**, In AD, mitochondrial ROS generation and inhibition of energy metabolism increase Aβ levels in cells and transgenic mice, and Aβ can interact with mitochondria and cause mitochondrial dysfunction. Aβ inhibits complex IV and α-ketoglutarate dehydrogenase (KGD), and binds Aβ-binding alcohol dehydrogenase (ABAD). Both KGD and ABAD produce ROS (white stars). Amyloid precursor protein (APP) may be targeted to the OMM and interfere with protein import. Mitochondria have also been reported to contain active γ-secretase complexes, which are involved in cleaving APP to form Aβ and contain presenilin 1, which increases the proteolytic activity of HTRA2 towards IAPs. AD patients have on average more somatic mutations in the mtDNA control region than control subjects. **b**, Complex I activity is decreased in PD, and inhibition of complex I by MPTP or rotenone causes parkinsonism. Mutations in mtDNA-encoded complex I subunits, 12S rRNA, and POLG also cause parkinsonism. Many genes associated with PD also implicate mitochondria in disease pathogenesis. α-Synuclein immunostaining is seen in degenerating mitochondria from mice overexpressing A53T α-synuclein. α-Synuclein overexpression impairs mitochondrial function and enhances the toxicity of MPTP. Parkin associates with the OMM and protects against cytochrome *c* release. It may also associate with mitochondrial-transcription-factor A (TFAM) and enhance mitochondrial biogenesis. When oxidized, DJ-1 translocates to mitochondria (IMS and matrix), downregulates the PTEN-tumour suppressor (not shown), and protects the cell from oxidative-stress-induced cell death. The mitochondrial kinase PINK1 protects against apoptosis, an effect that is reduced by PD-related mutations or kinase inactivation. Physical associations have been reported between DJ-1 and α-synuclein, DJ-1 and parkin, and DJ-1 and PINK1, and there is genetic evidence that DJ-1, PINK1 and parkin function sequentially in the same pathway. About 10% of the kinase LRRK2 is localized to mitochondria, and PD-related mutations augment its kinase activity. A mutation in HTRA2 was found in ~1% of sporadic PD patients. Overexpression of the mutant impaired normal HTRA2 protease activity, and HTRA2 knockout results in striatal degeneration and parkinsonism. **c**, Overexpression of mutant SOD1 in ALS impairs electron-transport-chain activities and decreases mitochondrial calcium-loading capacity. SOD1 has been localized to the OMM, IMS and matrix, and targeting of mutant SOD1 to mitochondria causes cytochrome *c* release and apoptosis. Mutant SOD1 promotes aberrant mitochondrial ROS production and forms aggregates that may clog the OMM protein importation machinery or bind and sequester the antiapoptotic protein Bcl-2. **d**, Complex II activity is decreased in the HD brain, and the complex-II inhibitor 3-nitropropionic acid induces striatal degeneration and movement disorder in rodents and primates. Overexpression of complex-II subunits reduces cell death in striatal neurons expressing mutant HTT. Mutant HTT associates with the OMM and increases sensitivity to calcium-induced cytochrome *c* release. Mutant HTT also translocates to the nucleus, where it binds and increases the level and transcriptional activity of p53. p53 activates the pro-apoptotic protein BAX, either directly or by increasing expression of BH3-only Bcl-2 family members NOXA and PUMA. In mice, knockout of *Pgc-1α* or a missense mutation in *Htra2* causes involuntary movements and striatal degeneration.

A53T mutant α -synuclein, degenerating mitochondria were immunostained for α -synuclein, raising the possibility that mutant α -synuclein might damage mitochondria directly⁵². Whereas overexpression of α -synuclein increases sensitivity to MPTP, α -synuclein-null mice are resistant to MPTP⁴¹ and other mitochondrial toxins such as malonate and 3-nitropropionic acid⁵³. Thus, α -synuclein seems to mediate the toxic effects of MPTP.

Mutations in parkin are associated with autosomal recessive juvenile PD. Parkin encodes a ubiquitin E3 ligase, and the primary abnormality, therefore, is in the ubiquitin–proteasome system. However, as above, there seem to be close interrelationships between the ubiquitin–proteasome system, oxidative stress and mitochondrial dysfunction. On one hand, parkin deficiency or mutations lead to oxidative stress and mitochondrial dysfunction. *Parkin*-null *Drosophila*⁵⁴ and mouse⁵⁵ strains exhibit mitochondrial impairment and increased oxidative stress, and leukocytes from individuals with parkin mutations have a selective impairment in complex-I activity⁵⁶. Parkin can associate with the outer mitochondrial membrane and prevent mitochondrial swelling, cytochrome *c* release and caspase activation, and this protective effect is abrogated by proteasome inhibitors and parkin mutations⁵⁷. Parkin has also been localized to mitochondria in proliferating cells, where it has been shown to associate with mitochondrial transcription factor A and to enhance mitochondrial biogenesis⁵⁸. On the other hand, mitochondrial dysfunction and oxidative stress can affect parkin function and exacerbate parkin mutations. S-nitrosylation of parkin, an oxidative modification, impairs its ubiquitin-ligase activity and compromises its protective function⁵⁹. Conversely, overexpression of glutathione S-transferase, which has a role in detoxifying products of oxidative damage, suppresses neurodegeneration in *Drosophila parkin* mutants⁶⁰.

Mutations in *DJ-1* are also associated with autosomal recessive juvenile PD⁶¹. *DJ-1* has been reported to interact with α -synuclein⁶², parkin⁶³ and PINK1 (ref. 64). The overall function of *DJ-1* seems to be to protect against cell death, especially that induced by oxidative stress. It can act as a redox sensor: oxidative stress causes a critical cysteine residue (C106) to be acidified, which leads to its relocalization to mitochondria. C106 mutations prevent this mitochondrial relocalization and impair the cell's response to oxidative stress and mitochondrial damage⁶⁵. *DJ-1* is a negative regulator of the PTEN tumour-suppressor protein, which promotes apoptosis by dephosphorylating phosphatidylinositol-3,4,5-trisphosphate, which is necessary for phosphatidylinositol 3OH-kinase-mediated activation of the cell-survival kinase Akt. *Dj-1* knockdown results in decreased phosphorylation of Akt⁶⁶, whereas *Dj-1* overexpression leads to Akt hyperphosphorylation and increased cell survival⁶⁷. *DJ-1*-deficient mice are hypersensitive to MPTP and oxidative stress⁶⁸. Moreover, in flies, *DJ-1* undergoes progressive oxidative inactivation with ageing, which, in turn, increases sensitivity to oxidative stress, and could provide one potential explanation for the age dependence of sporadic PD⁶⁹.

Mutations in *PINK1* represent a third form of autosomal recessive juvenile PD⁷⁰. *PINK1* is a kinase localized to mitochondria⁷¹, and, like *DJ-1*, seems to protect against cell death. Overexpression of wild-type *PINK1* prevents apoptosis under basal and staurosporine-induced conditions by decreasing cytochrome *c* release and caspase activation, and this effect is abrogated by PD-related mutations and a kinase-inactive mutation⁷². In *Drosophila*, *PINK1* deficiency causes mitochondrial pathology, increased sensitivity to paraquat and rotenone, and degeneration of flight muscles and dopaminergic neurons. This pathology resembles that of *parkin*-mutant flies, and can be rescued by overexpression of parkin, but not *DJ-1* (ref. 73). Thus, *PINK1* probably functions in the same pathway as parkin, with parkin downstream.

Mutations in *LRRK2* are the most common known cause of familial late-onset PD, and also account for 1–2% of sporadic late-onset PD cases. On the basis of its sequence, *LRRK2* is predicted to have a ROC-COR GTPase domain, a MAPK kinase domain and WD40 domains. Recently, West and colleagues showed that *LRRK2* is a kinase, that two disease-associated mutations, including the most common G2019S mutation, augment the kinase activity, and that ~10% of *LRRK2* is associated with mitochondria⁷⁴.

The gene most recently associated with PD is *HTRA2* (also known as *OMI*)⁷⁵. A G399S mutation was found in 4 of 414 individuals with sporadic PD and none of 313 controls. An additional polymorphism, A141S, was found in a heterozygous state in 3% of controls and 6.2% of PD subjects ($P=0.039$, odds ratio=2.15). As noted above (Fig. 1), *HTRA2* may be a protein quality control agent within mitochondria, and a pro-apoptotic factor when released into the cytosol from the mitochondrial intermembrane space through a mechanism mediated by the pro-apoptotic proteins BAX and BAK. Consistent with a role for *HTRA2* in maintaining mitochondrial function, homozygous *Htra2*-knockout mice develop striatal degeneration and parkinsonism⁷⁶. Expression in cultured cells of the G399S and A141S mutations found in people with PD impairs normal *HTRA2* protease activity, causes mitochondrial swelling, decreases mitochondrial membrane potential and increases staurosporine-induced cell death⁷⁵.

Mitochondria and amyotrophic lateral sclerosis

ALS is characterized clinically by progressive weakness, atrophy and spasticity of muscle tissue, reflecting the degeneration of upper and lower motor neurons in the cortex, brainstem and spinal cord. Approximately 90% of cases are sporadic (SALS) and 10% are familial (FALS). About 20% of familial cases are caused by mutations in Cu/Zn-superoxide dismutase (*SOD1*).

In both SALS and FALS, postmortem and biopsy samples from the spinal cord, nerves and muscles show abnormalities in mitochondrial structure, number and localization. Defects in activities of respiratory-chain complexes have also been detected in muscle and spinal cord. However, it is difficult to know from single snapshots of already symptomatic individuals whether mitochondria contribute to pathogenesis or are innocent bystanders. Thus, research on mitochondrial involvement in ALS has focused on expression of mutant *SOD1* in animal and cellular models of the disease.

Overexpressing the G93A *Sod1* mutation in transgenic mice causes impaired mitochondrial energy metabolism in the brain and spinal cord at disease onset⁷⁷. However, long before disease onset there is a decrease in the calcium-loading capacity in mitochondria from the brain and spinal cord, but not the liver, of mice overexpressing G93A or G85R mutant *Sod1* (ref. 78). In G93A *Sod1* mice there is a transient explosive increase in vacuolar mitochondrial degeneration just preceding motor-neuron death⁷⁹, suggesting that mitochondrial abnormalities trigger the onset of ALS.

Interestingly, *SOD1* immunoreactivity is concentrated inside vacuolated mitochondria⁸⁰. *SOD1* has traditionally been thought to be a cytoplasmic protein, but localization of a fraction of cellular *SOD1* to the mitochondrial outer membrane, intermembrane space and matrix has now been demonstrated^{81,82}. The localization of *SOD1* to mitochondria has been reported to occur only in affected tissues and to occur preferentially for mutant *SOD1* (ref. 83).

Interaction between *SOD1* and mitochondria suggests a number of mechanisms by which mitochondrial function and cell survival may be adversely affected. Mitochondrial targeting of mutant *SOD1* caused cytochrome *c* release and apoptosis, whereas targeting to the endoplasmic reticulum or nucleus did not cause cell death⁸⁴. Cleveland and colleagues suggest that mutant *SOD1* accumulates and aggregates in the outer mitochondrial membrane and clogs the protein importation machinery, eventually resulting in mitochondrial dysfunction⁸⁵. Mutant *SOD1* has been proposed to promote aberrant ROS production, and we found oxidative damage to mitochondrial lipids and proteins, accompanied by impaired respiration and ATP synthesis, in mice expressing mutant human *SOD1* (ref. 77). Mutant, but not wild-type, *SOD1* species bind to and aggregate with cytosolic heat-shock proteins⁸⁵ and mitochondrial Bcl-2 (ref. 86), rendering them unavailable for anti-apoptotic functions.

Mitochondria and Huntington's disease

HD is characterized clinically by chorea, psychiatric disturbances and dementia, and pathologically by loss of long projection neurons in the cortex and striatum. HD is inherited in an autosomal dominant

Table 2 | Mitochondrial involvement in less common neurodegenerative diseases

Disease	Clinical features	Protein	Function
Friedreich's ataxia	Ataxia and neuropathy due to degeneration of spinocerebellar tracts and dorsal-root ganglia; diabetes; cardiomyopathy Autosomal recessive	Frataxin	A mitochondrial iron chaperone that promotes the biogenesis of enzymes with Fe-S clusters and detoxifies excess iron. Frataxin deficiency causes iron accumulation and impairs the activity of Fe-S cluster-containing enzymes (complexes I and II, and aconitase). It may also perturb manganese balance and impair MnSOD activity
Hereditary spastic paraplegia (HSP)	Slowly progressive weakness and spasticity of the legs Autosomal recessive	SPG7*	An inner mitochondrial membrane m-AAA metalloprotease. It functions as a chaperone and is involved in the assembly of respiratory-chain complexes. SPG7 deficiency causes recessive HSP with mitochondrial myopathy
	Autosomal dominant	SPG13†	A mitochondrial protein chaperone
Neurodegeneration with brain iron accumulation (NBIA)	Progressive dementia, rigidity, involuntary movements, spasticity and retinal degeneration accompanied by iron deposition in globus pallidus and substantia nigra Autosomal recessive, paediatric onset	PANK2	PANK2 is localized to mitochondria and catalyses the first step in coenzyme A synthesis. PANK2 deficiency is the most common cause of NBIA, accounting for ~50% of cases
Optic atrophy type 1	Autosomal dominant optic neuropathy causing progressive visual loss	OPA1	OPA1 is a dynamin-related GTPase localized to mitochondria. It organizes the mitochondrial inner membrane and is necessary for maintaining cristae integrity. SiRNA knockdown of OPA1 causes fragmentation of the mitochondrial network, loss of mitochondrial membrane potential, disorganization of cristae, release of cytochrome c and activation of caspases

*Also known as paraplegin. †Also known as heat-shock protein 60. m-AAA, mitochondrial ATPases associated with diverse cellular activities; PANK2, pantothenate kinase 2; SPG, spastic paraplegia gene.

manner, and is due to expansion of a CAG trinucleotide repeat in the huntingtin (*HTT*) gene, which gives rise to an expanded polyglutamine stretch in the corresponding protein. The normal number of CAG (Q) repeats is less than 36; repeat numbers greater than 40 are associated with human disease.

Various lines of evidence demonstrate the involvement of mitochondrial dysfunction in HD. Nuclear magnetic resonance spectroscopy reveals increased lactate in the cortex and basal ganglia⁸⁷. Biochemical studies show decreased activities of complexes II and III of the electron-transport chain in the human HD brain⁸⁸. In striatal cells from mutant *Htt*-knock-in mouse embryos, mitochondrial respiration and ATP production are significantly impaired⁸⁹.

The mitochondrial dysfunction observed above is likely to be pathogenically important, because 3-nitropropionic acid and malonate — mitochondrial toxins that selectively inhibit succinate dehydrogenase and complex II — induce a clinical and pathological phenotype that closely resembles HD⁹⁰. Moreover, in striatal neurons expressing the first 171 amino acids of HTT with an insertion of 82 glutamines, overexpression of complex-II subunits restored complex-II activity and blocked mitochondrial dysfunction and cell death⁹¹.

There are several mechanisms by which the mutation could result in mitochondrial dysfunction. First, HTT may interact directly with mitochondria. In one study⁹², lymphoblast mitochondria from HD patients and brain mitochondria from YAC transgenic mice expressing HTT with 72 repeats were found to have lower membrane potentials and to depolarize at lower calcium loads than control mitochondria. Amino-terminal mutant HTT was identified on neuronal mitochondrial membranes with immunoelectron microscopy, and the incubation of normal mitochondria with mutant HTT reproduced the calcium-handling defect seen in HD patients and transgenic mice. In another study⁹³, subfractionation of mitochondria from a knock-in HD-mouse model showed HTT in association with the outer mitochondrial membrane. Mitochondria from the knock-in HD mouse were more sensitive to calcium-induced mitochondrial permeabilization and cytochrome *c* release, an effect that was mimicked by incubating normal mitochondria with mutant, but not wild-type, HTT.

Another mechanism through which mutant HTT could affect mitochondrial function is by altering transcription⁹⁴. HTT interacts with a number of transcription factors, including p53, CREB-binding protein and SP1 (for a review see ref. 95). p53 is a tumour suppressor known to regulate genes involved in mitochondrial function and oxidative stress. In response to genotoxic injury, p53 activates mitochondrial pathway apoptosis by increasing transcription of the pro-apoptotic BH3-only Bcl-2 family members such as PUMA⁹⁶. p53 can also translocate to mito-

chondria and directly activate BAX⁹⁷. In a recent study⁹⁸, mutant HTT bound p53 and increased p53 levels and transcriptional activity, leading to upregulation of downstream targets BAX and PUMA, and mitochondrial membrane depolarization. Pharmacological suppression or genetic deletion of p53 prevented HTT-induced mitochondrial depolarization, cytochrome oxidase deficiency and cytotoxicity. Moreover, expression of mutant HTT in a p53-null background diminished retinal degeneration in *Drosophila* and reversed behavioural abnormalities in mice.

Although they have not yet been directly related to human HD, two mitochondria-related proteins are associated with HD-like phenotypes in transgenic mice. PGC-1 α is a transcriptional coactivator that regulates mitochondrial biogenesis and metabolic pathways. *Pgc-1 α* -knockout mice exhibit impaired mitochondrial function, a hyperkinetic movement disorder and striatal degeneration — features that are all also observed in HD⁹⁹. HTRA2 is a serine protease that resides in the mitochondrial intermembrane space. In mice, a missense mutation in HTRA2 that reduces protease activity results in the Mnd2 (motor-neuron degeneration 2) phenotype, which shares many features of HD, including involuntary movements, abnormal postures and massive loss of striatal neurons¹⁰⁰.

Mitochondrial involvement in other, less common, neurodegenerative diseases is reviewed in Table 2.

Summary and future directions

Recent findings have greatly expanded our understanding of the role of mitochondria in the pathogenesis of neurodegenerative diseases. Mitochondrial-DNA mutations and oxidative stress contribute to ageing, the greatest risk factor for neurodegenerative diseases. Mitochondrial dysfunction and oxidative stress occur early in all major neurodegenerative diseases, and there is strong evidence that this dysfunction has a causal role in disease pathogenesis. Most impressively, specific interactions of disease-related proteins with mitochondria have recently been uncovered: APP, A β , presenilin, α -synuclein, parkin, DJ-1, PINK1, LRRK2, HTRA2, SOD1 and huntingtin have all been found within mitochondria, as discussed above. This explosion of findings raises further questions.

First, what is the basis for cell-type specificity in neurodegenerative disorders? For example, why does systemic overexpression of mutant SOD1 affect calcium handling in brain but not liver mitochondria⁷⁸? The propensity of mitochondrial disorders to affect the brain and muscles has thus far been explained by the different tissue requirements for mitochondrial function — brain and muscle tissues have high energy requirements. However, we speculate that mitochondrial differences between different cell types might be at least as important in this selectivity,

if not more so. Most mitochondrial proteins are encoded by nuclear genes, and it is to be expected that the distinct nuclear programmes required to produce different cell types should also produce different mitochondria. However, knowledge of mitochondrial biology in different cell types is extremely rudimentary. Further investigation will probably provide a more fundamental understanding of neurodegenerative diseases, because cell selectivity is a fundamental characteristic of these disorders.

Second, the complexity of mitochondrial ROS metabolism suggests that interventions such as the administration of one or a few antioxidants may be too simplistic. Indeed, such interventions have generally been, at best, modestly successful in clinical trials, despite abundant evidence for oxidative stress in disease pathogenesis. A more complete approach to antioxidant therapy would be to decrease ROS generation (for example, by expressing uncoupling proteins) and to upregulate the multilayered endogenous mitochondrial and intracellular antioxidant defence network. However, this will require a considerably better understanding of ROS biology than we have at present. It will also require the global 'ROS bad, antioxidants good' impression to be replaced with knowledge of the specific targets of ROS (such as PIN1 in AD).

Finally, the interaction of mitochondria with specific disease-related proteins opens exciting new possibilities for therapeutic targets. For example, in HD, reducing p53 protects against mutant huntingtin. However, in the case of AD, PD and ALS, genetic forms of disease account for only a small percentage of cases, and the relevance of mitochondrial interactions with these proteins must be determined for sporadic cases.

Knowledge of neurodegenerative diseases has advanced rapidly in the last few years, and the field holds great promise for furthering our understanding and the eventual treatment of these devastating illnesses. ■

- Daniel, N. N. & Korsmeyer, S. J. Cell death: critical control points. *Cell* **116**, 205–219 (2004).
- Corral-Debrinski, M. *et al.* Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age. *Nature Genet.* **2**, 324–329 (1992).
- Zhang, J. *et al.* Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. *Proc. Natl Acad. Sci. USA* **100**, 1116–1121 (2003).
- Chinnery, P. F. *et al.* Point mutations of the mtDNA control region in normal and neurodegenerative human brains. *Am. J. Hum. Genet.* **68**, 529–532 (2001).
- Simon, D. K. *et al.* Low mutational burden of individual acquired mitochondrial DNA mutations in brain. *Genomics* **73**, 113–116 (2001).
- Lin, M. T., Simon, D. K., Ahn, C. H., Kim, L. M. & Beal, M. F. High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. *Hum. Mol. Genet.* **11**, 133–145 (2002).
- Jazin, E. E., Cavellier, L., Eriksson, I., Orelund, L. & Gyllenstein, U. Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. *Proc. Natl Acad. Sci. USA* **93**, 12382–12387 (1996).
- Bender, A. *et al.* High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nature Genet.* **38**, 515–517 (2006).
- Kraytsberg, Y. *et al.* Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nature Genet.* **38**, 518–520 (2006).
- Trifunovic, A. Mitochondrial DNA and ageing. *Biochim. Biophys. Acta* **1757**, 611–617 (2006).
- Andreyev, A. Y., Kushnareva, Y. E. & Starkov, A. A. Mitochondrial metabolism of reactive oxygen species. *Biochemistry (Mosc.)* **70**, 200–214 (2005).
- Sun, J., Folk, D., Bradley, T. J. & Tower, J. Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster*. *Genetics* **161**, 661–672 (2002).
- Ruan, H. *et al.* High-quality life extension by the enzyme peptide methionine sulfoxide reductase. *Proc. Natl Acad. Sci. USA* **99**, 2748–2753 (2002).
- Schriner, S. E. *et al.* Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science* **308**, 1909–1911 (2005).
- Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
- Nunomura, A. *et al.* Oxidative damage is the earliest event in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* **60**, 759–767 (2001).
- Pratico, D., Uryu, K., Leight, S., Trojanowski, J. Q. & Lee, V. M. Increased lipid peroxidation precedes amyloid plaque formation in an animal model of Alzheimer amyloidosis. *J. Neurosci.* **21**, 4183–4187 (2001).
- Reddy, P. H. *et al.* Gene expression profiles of transcripts in amyloid precursor protein transgenic mice: up-regulation of mitochondrial metabolism and apoptotic genes is an early cellular change in Alzheimer's disease. *Hum. Mol. Genet.* **13**, 1225–1240 (2004).
- Ohayagi, Y. *et al.* Selective increase in cellular A β 42 is related to apoptosis but not necrosis. *Neuroreport* **11**, 167–171 (2000).
- Busciglio, J. *et al.* Altered metabolism of the amyloid β precursor protein is associated with mitochondrial dysfunction in Down's syndrome. *Neuron* **33**, 677–688 (2002).
- Li, F. *et al.* Increased plaque burden in brains of APP mutant MnSOD heterozygous knockout mice. *J. Neurochem.* **89**, 1308–1312 (2004).
- Velliquette, R. A., O'Connor, T. & Vassar, R. Energy inhibition elevates β -secretase levels and activity and is potentially amyloidogenic in APP transgenic mice: possible early events in Alzheimer's disease pathogenesis. *J. Neurosci.* **25**, 10874–10883 (2005).
- Tamagno, E. *et al.* β -Site APP cleaving enzyme up-regulation induced by 4-hydroxynonenal is mediated by stress-activated protein kinases pathways. *J. Neurochem.* **92**, 628–636 (2005).
- Lovell, M. A., Xiong, S., Xie, C., Davies, P. & Markesbery, W. R. Induction of hyperphosphorylated tau in primary rat cortical neuron cultures mediated by oxidative stress and glycogen synthase kinase-3. *J. Alzheimers Dis.* **6**, 659–671 (2004).
- Sultana, R. *et al.* Oxidative modification and down-regulation of Pin1 in Alzheimer's disease hippocampus: a redox proteomics analysis. *Neurobiol. Aging* **27**, 918–925 (2006).
- Pastorino, L. *et al.* The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid- β production. *Nature* **440**, 528–534 (2006).
- Liou, Y. C. *et al.* Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. *Nature* **424**, 556–561 (2003).
- Swerdlow, R. H. *et al.* Cybrids in Alzheimer's disease: a cellular model of the disease? *Neurology* **49**, 918–925 (1997).
- Elson, J. L. *et al.* Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Hum. Genet.* **119**, 241–254 (2006).
- Coskun, P. E., Beal, M. F. & Wallace, D. C. Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication. *Proc. Natl Acad. Sci. USA* **101**, 10726–10731 (2004).
- Anandatheerthavarada, H. K., Biswas, G., Robin, M. A. & Avadhani, N. G. Mitochondrial targeting and a novel transmembrane arrest of Alzheimer's amyloid precursor protein impairs mitochondrial function in neuronal cells. *J. Cell Biol.* **161**, 41–54 (2003).
- Lustbader, J. W. *et al.* A β AD directly links A β to mitochondrial toxicity in Alzheimer's disease. *Science* **304**, 448–452 (2004).
- Crouch, P. J. *et al.* Copper-dependent inhibition of human cytochrome c oxidase by a dimeric conformer of amyloid- β_{1-42} . *J. Neurosci.* **25**, 672–679 (2005).
- Manczak, M. *et al.* Mitochondria are a direct site of A β accumulation in Alzheimer's disease neurons: implications for free radical generation and oxidative damage in disease progression. *Hum. Mol. Genet.* **15**, 1437–1449 (2006).
- Casley, C. S., Canevari, L., Land, J. M., Clark, J. B. & Sharpe, M. A. β -amyloid inhibits integrated mitochondrial respiration and key enzyme activities. *J. Neurochem.* **80**, 91–100 (2002).
- Gibson, G. E. *et al.* Reduced activities of thiamine-dependent enzymes in the brains and peripheral tissues of patients with Alzheimer's disease. *Arch. Neurol.* **45**, 836–840 (1988).
- Parker, W. D., Filley, C. M. & Parks, J. K. Cytochrome oxidase deficiency in Alzheimer's disease. *Neurology* **40**, 1302–1303 (1990).
- Park, H. J., Seong, Y. M., Choi, J. Y., Kang, S. & Rhim, H. Alzheimer's disease-associated amyloid β interacts with the human serine protease HtrA2/Omi. *Neurosci. Lett.* **357**, 63–67 (2004).
- Hansson, C. A. *et al.* Nicastrin, presenilin, APH-1, and PEN-2 form active γ -secretase complexes in mitochondria. *J. Biol. Chem.* **279**, 51654–51660 (2004).
- Betarbet, R. *et al.* Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neurosci.* **3**, 1301–1306 (2000).
- Fornai, F. *et al.* Parkinson-like syndrome induced by continuous MPTP infusion: convergent roles of the ubiquitin-proteasome system and α -synuclein. *Proc. Natl Acad. Sci. USA* **102**, 3413–3418 (2005).
- Sherer, T. B. *et al.* Mechanism of toxicity in rotenone models of Parkinson's disease. *J. Neurosci.* **23**, 10756–10764 (2003).
- Schapira, A. H. *et al.* Mitochondrial complex I deficiency in Parkinson's disease. *Lancet* **i**, 1269 (1989).
- Simon, D. K. *et al.* Familial multisystem degeneration with parkinsonism associated with the 11778 mitochondrial DNA mutation. *Neurology* **53**, 1787–1793 (1999).
- Luoma, P. *et al.* Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: clinical and molecular genetic study. *Lancet* **364**, 875–882 (2004).
- Vives-Bauza, C. *et al.* Sequence analysis of the entire mitochondrial genome in Parkinson's disease. *Biochem. Biophys. Res. Commun.* **290**, 1593–1601 (2002).
- Simon, D. K. *et al.* Somatic mitochondrial DNA mutations in cortex and substantia nigra in aging and Parkinson's disease. *Neurobiol. Aging* **25**, 71–81 (2004).
- Pyle, A. *et al.* Mitochondrial DNA haplogroup cluster UK1T reduces the risk of PD. *Ann. Neurol.* **57**, 564–567 (2005).
- Tanaka, M. Mitochondrial genotypes and cytochrome b variants associated with longevity or Parkinson's disease. *J. Neurol.* **249** (Suppl. 2), 1111–1118 (2002).
- Wallace, D. C. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu. Rev. Genet.* **39**, 359–407 (2005).
- Song, D. D., Shults, C. W., Sisk, A., Rockenstein, E. & Masliah, E. Enhanced substantia nigra mitochondrial pathology in human α -synuclein transgenic mice after treatment with MPTP. *Exp. Neurol.* **186**, 158–172 (2004).
- Martin, L. J. *et al.* Parkinson's disease α -synuclein transgenic mice develop neuronal mitochondrial degeneration and cell death. *J. Neurosci.* **26**, 41–50 (2006).
- Klivenyi, P. *et al.* Mice lacking α -synuclein are resistant to mitochondrial toxins. *Neurobiol. Dis.* **21**, 541–548 (2006).
- Pesah, Y. *et al.* *Drosophila* parkin mutants have decreased mass and cell size and increased sensitivity to oxygen radical stress. *Development* **131**, 2183–2194 (2004).
- Palacio, J. J. *et al.* Mitochondrial dysfunction and oxidative damage in parkin-deficient mice. *J. Biol. Chem.* **279**, 18614–18622 (2004).
- Muftuoglu, M. *et al.* Mitochondrial complex I and IV activities in leukocytes from patients with parkin mutations. *Mov. Disord.* **19**, 544–548 (2004).
- Darios, F. *et al.* Parkin prevents mitochondrial swelling and cytochrome c release in mitochondria-dependent cell death. *Hum. Mol. Genet.* **12**, 517–526 (2003).
- Kuroda, Y. *et al.* Parkin enhances mitochondrial biogenesis in proliferating cells. *Hum. Mol. Genet.* **15**, 883–895 (2006).
- Chung, K. K. *et al.* S-nitrosylation of parkin regulates ubiquitination and compromises parkin's protective function. *Science* **304**, 1328–1331 (2004).

60. Whitworth, A. J. *et al.* Increased glutathione S-transferase activity rescues dopaminergic neuron loss in a *Drosophila* model of Parkinson's disease. *Proc. Natl Acad. Sci. USA* **102**, 8024–8029 (2005).
61. Bonifati, V. *et al.* Mutations in the *DJ-1* gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256–259 (2003).
62. Meulener, M. C. *et al.* DJ-1 is present in a large molecular complex in human brain tissue and interacts with α -synuclein. *J. Neurochem.* **93**, 1524–1532 (2005).
63. Moore, D. J. *et al.* Association of DJ-1 and parkin mediated by pathogenic DJ-1 mutations and oxidative stress. *Hum. Mol. Genet.* **14**, 71–84 (2005).
64. Tang, B. *et al.* Association of PINK1 and DJ-1 confers digenic inheritance of early-onset Parkinson's disease. *Hum. Mol. Genet.* **15**, 1816–1825 (2006).
65. Canet-Aviles, R. M. *et al.* The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc. Natl Acad. Sci. USA* **101**, 9103–9108 (2004).
66. Yang, Y. *et al.* Inactivation of *Drosophila* DJ-1 leads to impairments of oxidative stress response and phosphatidylinositol 3-kinase/Akt signaling. *Proc. Natl Acad. Sci. USA* **102**, 13670–13675 (2005).
67. Kim, R. H. *et al.* DJ-1, a novel regulator of the tumor suppressor PTEN. *Cancer Cell* **7**, 263–273 (2005).
68. Kim, R. H. *et al.* Hypersensitivity of DJ-1-deficient mice to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) and oxidative stress. *Proc. Natl Acad. Sci. USA* **102**, 5215–5220 (2005).
69. Meulener, M. C., Xu, K., Thompson, L., Ischiropoulos, H. & Bonini, N. M. Mutational analysis of DJ-1 in *Drosophila* implicates functional inactivation by oxidative damage and aging. *Proc. Natl Acad. Sci. USA* **103**, 12517–12522 (2006).
70. Valente, E. M. *et al.* Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. *Science* **304**, 1158–1160 (2004).
71. Silvestri, L. *et al.* Mitochondrial import and enzymatic activity of PINK1 mutants associated to recessive parkinsonism. *Hum. Mol. Genet.* **14**, 3477–3492 (2005).
72. Petit, A. *et al.* Wild-type PINK1 prevents basal and induced neuronal apoptosis, a protective effect abrogated by Parkinson disease-related mutations. *J. Biol. Chem.* **280**, 34025–34032 (2005).
73. Yang, Y. *et al.* Mitochondrial pathology and muscle and dopaminergic neuron degeneration caused by inactivation of *Drosophila* PINK1 is rescued by Parkin. *Proc. Natl Acad. Sci. USA* **103**, 10793–10798 (2006).
74. West, A. B. *et al.* Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl Acad. Sci. USA* **102**, 16842–16847 (2005).
75. Strauss, K. M. *et al.* Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Hum. Mol. Genet.* **14**, 2099–2111 (2005).
76. Martins, L. M. *et al.* Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice. *Mol. Cell. Biol.* **24**, 9848–9862 (2004).
77. Mattiazzi, M. *et al.* Mutated human SOD1 causes dysfunction of oxidative phosphorylation in mitochondria of transgenic mice. *J. Biol. Chem.* **277**, 29626–29633 (2002).
78. Damiano, M. *et al.* Neural mitochondrial Ca^{2+} capacity impairment precedes the onset of motor symptoms in G93A Cu/Zn-superoxide dismutase mutant mice. *J. Neurochem.* **96**, 1349–1361 (2006).
79. Kong, J. & Xu, Z. Massive mitochondrial degeneration in motor neurons triggers the onset of amyotrophic lateral sclerosis in mice expressing a mutant SOD1. *J. Neurosci.* **18**, 3241–3250 (1998).
80. Jaarsma, D. *et al.* CuZn superoxide dismutase (SOD1) accumulates in vacuolated mitochondria in transgenic mice expressing amyotrophic lateral sclerosis-linked SOD1 mutations. *Acta Neuropathol. (Berl.)* **102**, 293–305 (2001).
81. Higgins, C. M., Jung, C. & Xu, Z. ALS-associated mutant SOD1G93A causes mitochondrial vacuolation by expansion of the intermembrane space and by involvement of SOD1 aggregation and peroxisomes. *BMC Neurosci.* **4**, 16 (2003).
82. Vijayvergiya, C., Beal, M. F., Buck, J. & Manfredi, G. Mutant superoxide dismutase 1 forms aggregates in the brain mitochondrial matrix of amyotrophic lateral sclerosis mice. *J. Neurosci.* **25**, 2463–2470 (2005).
83. Liu, J. *et al.* Toxicity of familial ALS-linked SOD1 mutants from selective recruitment to spinal mitochondria. *Neuron* **43**, 5–17 (2004).
84. Takeuchi, H., Kobayashi, Y., Ishigaki, S., Doyu, M. & Sobue, G. Mitochondrial localization of mutant superoxide dismutase 1 triggers caspase-dependent cell death in a cellular model of familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **277**, 50966–50972 (2002).
85. Okado-Matsumoto, A. & Fridovich, I. Amyotrophic lateral sclerosis: a proposed mechanism. *Proc. Natl Acad. Sci. USA* **99**, 9010–9014 (2002).
86. Pasinelli, P. *et al.* Amyotrophic lateral sclerosis-associated SOD1 mutant proteins bind and aggregate with Bcl-2 in spinal cord mitochondria. *Neuron* **43**, 19–30 (2004).
87. Jenkins, B. G., Koroshetz, W. J., Beal, M. F. & Rosen, B. R. Evidence for impairment of energy metabolism *in vivo* in Huntington's disease using localized ^1H NMR spectroscopy. *Neurology* **43**, 2689–2695 (1993).
88. Gu, M. *et al.* Mitochondrial defect in Huntington's disease caudate nucleus. *Ann. Neurol.* **39**, 385–389 (1996).
89. Milakovic, T. & Johnson, G. V. Mitochondrial respiration and ATP production are significantly impaired in striatal cells expressing mutant huntingtin. *J. Biol. Chem.* **280**, 30773–30782 (2005).
90. Brouillet, E. *et al.* Chronic mitochondrial energy impairment produces selective striatal degeneration and abnormal choreiform movements in primates. *Proc. Natl Acad. Sci. USA* **92**, 7105–7109 (1995).
91. Benchoua, A. *et al.* Involvement of mitochondrial complex II defects in neuronal death produced by N-terminus fragment of mutated huntingtin. *Mol. Biol. Cell* **17**, 1652–1663 (2006).
92. Panov, A. V. *et al.* Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. *Nature Neurosci.* **5**, 731–736 (2002).
93. Choo, Y. S., Johnson, G. V., MacDonald, M., Detloff, P. J. & Lesort, M. Mutant huntingtin directly increases susceptibility of mitochondria to the calcium-induced permeability transition and cytochrome *c* release. *Hum. Mol. Genet.* **13**, 1407–1420 (2004).
94. Luthi-Carter, R. & Cha, J.-H. J. Mechanisms of transcription dysregulation in Huntington's disease. *Clin. Neurosci. Res.* **3**, 165–177 (2003).
95. Sugars, K. L. & Rubinstein, D. C. Transcriptional abnormalities in Huntington disease. *Trends Genet.* **19**, 233–238 (2003).
96. Yu, J., Zhang, L., Hwang, P. M., Kinzler, K. W. & Vogelstein, B. PUMA induces the rapid apoptosis of colorectal cancer cells. *Mol. Cell* **7**, 673–682 (2001).
97. Chipuk, J. E. *et al.* Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science* **303**, 1010–1014 (2004).
98. Bae, B. I. *et al.* p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. *Neuron* **47**, 29–41 (2005).
99. Lin, J. *et al.* Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1 α null mice. *Cell* **119**, 121–135 (2004).
100. Rathke-Hartlieb, S. *et al.* Progressive loss of striatal neurons causes motor dysfunction in MND2 mutant mice and is not prevented by Bcl-2. *Exp. Neurol.* **175**, 87–97 (2002).

Acknowledgments This work was supported by grants from the National Institutes of Health and the American Federation for Aging Research/Beeson Program. We apologize to our many colleagues whose work we were unable to cite due to space limitations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to M.F.B. (fbeal@med.cornell.edu).

Cell death in the nervous system

Dale E. Bredesen^{1,2}, Rammohan V. Rao¹ & Patrick Mehlen^{1,3}

Neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease trigger neuronal cell death through endogenous suicide pathways. Surprisingly, although the cell death itself may occur relatively late in the course of the degenerative process, the mediators of the underlying cell-death pathways have shown promise as potential therapeutic targets.

Thank Heaven! The crisis — the danger, is past, and the lingering illness, is over at last — and the fever called "Living" is conquered at last.

from *For Annie*, by Edgar Allan Poe

Programmed cell death (PCD) has a critical role in the development of the nervous system, and both anti-PCD and pro-PCD modulators feature prominently in the establishment of neural architecture. It has been 100 years since the first description of developmental neuronal-cell death¹, and more than 50 years since Levi-Montalcini showed that such physiological cell death is inhibited by soluble factors such as nerve growth factor². Dysregulation of cell-death programmes features in developmental and neoplastic disorders of the nervous system, and there is increasing evidence to suggest that such dysregulation may also occur in neurodegenerative, infectious, traumatic, ischaemic, metabolic and demyelinating disorders.

In 1964, Lockshin and his colleagues introduced the term programmed cell death to describe the apparently predetermined pattern by which specific cells die during insect development³. In 1966, it was shown that this process requires protein synthesis, at least in some cases², indicating that it is the result of an active cellular suicide process. Then, in 1972, John Kerr and his colleagues coined the term apoptosis to describe a morphologically relatively uniform set of cell deaths seen in many different situations, from development to insult response to cell turnover⁴.

Although PCD has often been equated with apoptosis, non-apoptotic forms also exist^{5–9}, and neurodegenerative conditions such as Huntington's disease, amyotrophic lateral sclerosis (ALS) and ischaemia show cell deaths that do not fulfil the criteria for apoptosis⁷.

Classical developmental studies support the view that at least three different forms of PCD are distinguishable (Table 1): type I, also known as nuclear or apoptotic; type II, also known as autophagic^{3,5,8}; and type III, also known as cytoplasmic^{5,6,8,10}. These occur reproducibly in specific nuclei and with specific frequencies, at particular times of nervous-system development. But these cell-death pathways may also be activated by various insults, such as DNA damage or the accumulation of misfolded proteins.

Neurodegenerative diseases are associated with a number of insults that may trigger PCD: misfolded proteins, reactive oxygen and nitrogen species, mitochondrial-complex inhibition, calcium entry, excitotoxicity, trophic-factor withdrawal, and death-receptor activation to name a few. In some cases, however, deaths occur that do not fit neatly into any of the three classes of PCD, and these more controversial forms of death are also discussed below.

Temporal studies of neurodegenerative models suggest, however, that PCD may be a relatively late event in the neurodegenerative process, and that death is preceded by early functional alterations (for example, electrophysiological deficits and cellular-stress-pathway activation) and microanatomical deficits (such as neurite retraction and synapse loss; see page 768). Surprisingly, then, various approaches aimed at inhibiting PCD have led to improved outcomes in neurodegenerative models, indicating that these pathways could have an important role in neurodegenerative diseases. Furthermore, recent studies have suggested that death in the nervous system may trigger stem-cell proliferation and survival, and so the work on cell death pathways — the subject of this review — offers many potential points of entry into the therapeutics of neurodegenerative disease states.

Classical apoptosis

Apoptosis (from the Greek, 'falling away'), also referred to as nuclear or type I PCD, is the best-characterized type of PCD (Box 1). Morphologically, cells typically round up, form blebs, undergo zeiosis (an appearance of boiling due to rapid bleb formation), chromatin condensation, nuclear fragmentation and the budding off of apoptotic bodies. Phosphatidylserine, which is usually located at the plasma membrane and faces inwards on live cells, now faces both inwards and outwards¹¹. These morphological and histochemical changes are largely the result of activation of a set of cell-suicide cysteine proteases, the caspases^{12,13}.

The biochemical activation of classical apoptosis occurs through two main pathways (Box 1). These are the extrinsic pathway, which originates through the activation of cell-surface death receptors such as Fas, and results in the activation of caspase-8 or -10 (ref. 14), and the intrinsic pathway, which originates from mitochondrial release of cytochrome *c* and associated activation of caspase-9. A third, less well-characterized pathway — essentially a second intrinsic pathway — originates from the endoplasmic reticulum (ER) and also results in the activation of caspase-9 (refs 15–17). Other organelles, such as the nucleus and Golgi apparatus, have damage sensors that link to apoptotic pathways¹⁸.

Various biochemical responders, both physiological and pathological, act on the intrinsic pathway of apoptosis. Whether triggered by DNA damage or by sensors associated with other organelles, the typical outcome is that the balance between pro-apoptotic members of the Bcl-2 family and anti-apoptotic members is shifted toward the pro-apoptotic members (Box 1), which leads (although not invariably) to the cell's demise. But both the intrinsic and the extrinsic apoptotic pathways ultimately rely on the activation of caspases for death to ensue.

Caspases are cysteine-aspartyl-specific proteases that cleave with remarkable specificity at a small subset of aspartic acid residues. There are two types of apoptotic caspase: initiators and effectors. The initiator

¹Buck Institute for Age Research, 8001 Redwood Boulevard, Novato, California 94945, USA. ²University of California, San Francisco, 450 Parnassus, San Francisco, California 94143, USA.

³Apoptosis, Cancer and Development Laboratory, Université de Lyon, Centre Léon Bérard, 28 Rue Laennec, 69008 Lyon, France.

caspases cleave inactive forms of effector caspases, thereby activating them; effector caspases (for example, caspase-3 and 7) in turn cleave other protein substrates in the cell, resulting in the apoptotic process.

Both type of caspase are relatively inactive (the effectors more so than the initiators) when first synthesized, but they differ markedly in terms of their activation. Initiator caspases (caspase-8, 9 and 10) exist as monomers and bind to other proteins by means of what is known as the caspase activation and recruitment domain (CARD) in caspase-9 and a death effector domain (DED) in caspase-8 and 10. This protein–protein interaction results in dimerization of the caspases, which leads to their activation. Contrary to earlier models, cleavage of initiator caspases is neither required nor sufficient for activation¹⁹. However, the effector caspases exist as dimers in the cell and are activated by cleavage rather than induced proximity. Cleavage produces a tetramer with two large subunits and two small subunits with a substrate specificity that differs from initiator caspases.

The caspase substrates — the number of which is unknown but is probably somewhere between 0.5% and 5% of proteins — contribute to the apoptotic phenotype in various ways, such as by activation of proteolytic cascades, inactivation of repair, DNA cleavage, mitochondrial permeabilization and initiation of the process of phagytosis to clear up the dying cells, apoptotic bodies and debris.

Origami meets apoptosis

Misfolded proteins are constantly being produced, and for some proteins such misfolded species represent a significant fraction of the overall output. Misfolded proteins trigger a protective stress response, known as the unfolded-protein response (UPR; Fig. 1). Although this response may put off a cellular catastrophe for a short time, prolonged ER stress and UPR activation completely overwhelm the cellular protective mechanism, ultimately resulting in the activation of suicide pathways^{20–23} (Box 2). Moreover, misfolded proteins also aggregate as oligomers and higher-order multimers, both of which may interact with critical cellular targets such as chaperones and transcription factors, among others.

These pathways are of particular interest in neurodegenerative disease studies because they are implicated in all the main examples of such diseases²⁴. Disorders such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), ALS and prion-protein diseases all share one common feature: accumulation and aggregation of misfolded proteins^{20,21} (see pages 774 and 803).

Mediators of cell death induced by misfolded proteins have recently been identified (Box 2). Bcl-2 family proteins also seem to have a key role in the cellular-suicide decision process²⁵, and in the communication between the ER and the mitochondria. BAX/BAK double knockout cells demonstrate no caspase activation following ER stress, indicating that these are required mediators²⁶. BIK may function to activate BAX and BAK in this pathway, whereas BI-1 inhibits BAX activation and translocation to the ER²⁷. Other Bcl-2 family proteins are also involved, such as PUMA and NOXA, as well as p53 (ref. 28).

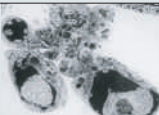
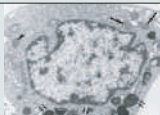
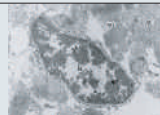
Rather than true protein misfolding, prion proteins may trigger PCD when they are in an alternative, physiologically relevant conformation²⁹. The prion protein exists in three different topologies: a secreted form, a transmembrane form in which the amino terminus is extracellular (NTM), and a transmembrane form in which the carboxy terminus is extracellular (CTM). The latter form is pro-apoptotic and associated with neurodegeneration *in vivo*, whereas the secreted form is anti-apoptotic. It remains to be seen whether other proteins have similar features.

Autophagic cell death

Desperate times call for desperate measures, and cells have a host of protective stress responses, most of which switch into execution mode during prolonged activation. During starvation, cells may only be able to survive by a process known as autophagy (from the Greek, 'self eating'). Autophagy occurs in diverse organisms and is subdivided into macroautophagy, microautophagy and chaperone-mediated autophagy. It complements the proteasomal pathway in that long-lived proteins, protein aggregates, and organelles are degraded by this regulated lysosomal pathway of degradation (see page 780). Targets for degradation — for example, damaged mitochondria or aggregates of misfolded proteins — are encircled by a process that is essentially an intracellular form of phagocytosis. The newly membrane-delimited structure — an autophagosome — then fuses with a lysosome, resulting in the degradation of the contents of the autophagosome. The molecular details of this process have been best characterized in yeast, in which a number of ATG (autophagy) genes have been identified, most of which have clear orthologues in higher eukaryotes.

Because the degradation of molecules and organelles by autophagy results in the production of energy and amino acids for protein synthesis, it is a cellular protective pathway that, although constitutively active at a low level, can be markedly upregulated by nutrient starvation.

Table 1 | Characteristics of dying cells

Characteristics	Established forms of cell death Apoptosis	Autophagic	Atypical forms of cell death Paraptosis	Calcium-mediated	AIF/PARP-dependent	Oncosis
Morphology	 Chromatin condensation, nuclear fragmentation, apoptotic bodies	 Autophagic vacuoles	 ER swelling, mitochondrial swelling	 Membrane whorls	 Mild chromatin condensation	 Cellular swelling
Triggers	Include death receptors, trophic-factor withdrawal, DNA damage, viral infections	Serum, amino-acid starvation, protein aggregates	Trophotoxicity	Calcium entry, <i>C. elegans deg</i> mutants	DNA damage, glutamate, nitric oxide	Ischaemia, excitotoxicity
Mediators	Caspases, BH1-3, BH3 proteins	JNK1? MKK7? ATG orthologues	ERK2, NUR77	Calpains, cathepsins	PARP, AIF	JNK
Inhibitors	Caspase inhibitors, BH1-4 proteins	JNK inhibitors?	U0126 (MEK), DN NUR77	Calreticulin, Some calpain inhibitors?	PARP inhibitors	JNK inhibitors
Examples	Type I PCD, nuclear PCD	Type II PCD	Type III PCD, cytoplasmic PCD	<i>C. elegans deg</i> mutants	Some excitotoxic PCD	Ischaemic PCD

Note the difference in morphology present in each form, as well as the differences in biochemical mediators, inducers, and inhibitors. At present, only apoptosis and autophagic PCD are generally accepted as being legitimate forms of PCD; however, ongoing research should reveal which of the additional candidates represent novel pathways of PCD. DN, dominant-negative; DEG, degeneration; U0126, a mitogen-activated protein kinase inhibitor. (Images reproduced, with permission, from refs 58–62.)

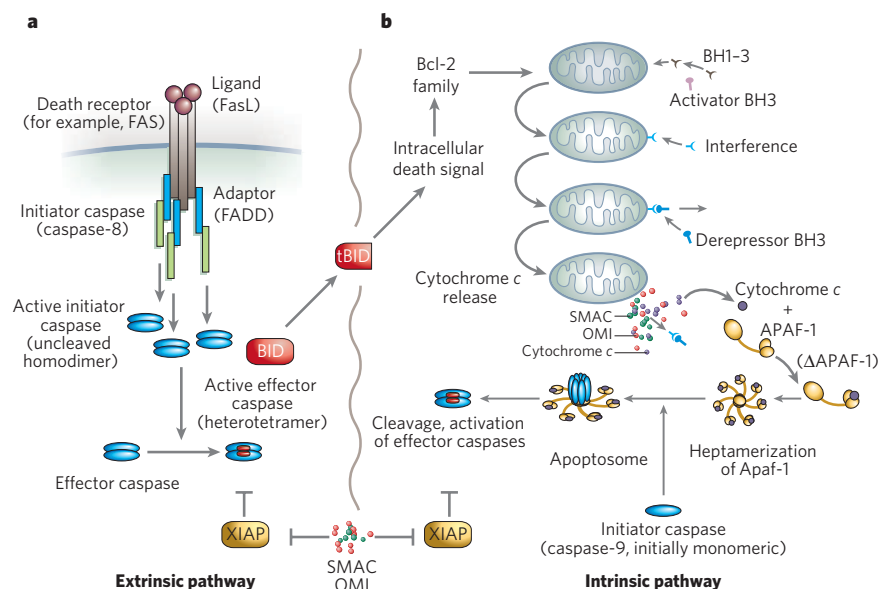
Box 1 | Intrinsic and extrinsic apoptotic pathways

In the best-characterized example of the extrinsic pathway (panel **a**) Fas is bound by the trimeric Fas ligand (FasL). This results in the recruitment of FADD through Fas's death domain (DD) and caspase-8 through FADD's death effector domain (DED)⁶³. The proximity of the initiator caspase, which results from its recruitment to the trimeric Fas–FADD complex, leads to its activation, and the subsequent activation of effector caspases (such as caspase-3 and 7) by cleavage. Furthermore, the extrinsic pathway interacts with the intrinsic pathway via caspase-8 cleavage of BID to produce tBID.

Whereas the extrinsic pathway uses caspase-8 (or caspase-10) to initiate cell death, the intrinsic pathway relies on caspase-9 (panel **b**). The propensity of cells to undergo apoptosis, known as the apoptat, is largely determined by the balance between anti-apoptotic and pro-apoptotic members of the Bcl-2 family of proteins. There are three types of pro-apoptotic Bcl-2-family protein. First, the multidomain proteins, such as BAX and BAK, have Bcl-2 homology domains 1–3 (BH1–3), which can permeabilize mitochondrial outer membranes⁶⁴. Second, the BH3-only proteins, such as BIM and tBID, activate BAX and BAK and may participate in pore formation. Third, the BH3-only derepressors such as PUMA, NOXA and BAD sequester the anti-apoptotic Bcl-2 and Bcl-X_L (and related proteins with BH1–4 domains), freeing up BH1–3 proteins to permeabilize the mitochondria. This causes the release of various pro-apoptotic mitochondrial proteins, such as cytochrome c and SMAC (also known as DIABLO), into the cytosol. Other proteins that are unrelated to Bcl-2 can also influence the actions of the Bcl-2 family members^{65–67}.

After release from the mitochondria, cytochrome c induces conformational change and heptamerization of the cytosolic protein APAF-1. The heptamer binds caspase-9, which results in its activation and the cleavage of effector caspases such as caspase-3 and 7. These active caspases can still be held in check by inhibitor of apoptosis proteins (IAPs) such as XIAP⁶⁸, which may function as direct inhibitors of caspase activity and as E3 ligases that mediate caspase degradation⁶⁹. This IAP-mediated block may be released by proteins such as SMAC^{70,71} and OMI (also known as HTRA2)⁷², which are also released from the mitochondria.

Fission and fusion of mitochondria^{73,74} also seem to function in cell death. Fission is mediated by DRP1 and FIS1, and inhibition of these proteins blocks the induction of apoptosis by staurosporine⁷⁴. Pro-apoptotic proteins may be recruited to the mitochondria by DRP1, and mitochondrial remodelling may enhance the release of cytochrome c and other mitochondrial proteins. This could have special relevance to PCD during neurodegeneration, because mutations in the mitochondrial fusion mediator OPA1 are associated with optic atrophy^{75,76}. Ongoing work should clarify the roles of mitochondrial fission and fusion, and their mediators, in PCD and neurodegeneration.



Nutrient withdrawal inactivates TOR (target of rapamycin), activating an ATG complex³⁰. In a second step, vesicle nucleation occurs, and then, in a third step, vesicle expansion occurs, followed finally by the recycling of ATG proteins. The importance of this pathway *in vivo* has been illustrated by *Atg7* conditionally null mice³¹, which show various cellular abnormalities such as ubiquitin-positive aggregates and apparently damaged mitochondria. Furthermore, mice deficient in autophagy due to knockout of beclin 1 are nonviable³².

Although the roles of the autophagic process in protein and organelle degradation, and in cellular protection during nutrient starvation, are well accepted, the role of autophagy in PCD is more controversial^{30,32,33}. This is in part because the term 'autophagic cell death' has been used for two potentially distinct observations: cell death associated with autophagy and cell death requiring autophagy. Most examples of autophagic cell death represent the former. However, increasing evidence suggests that the autophagic process is required for at least some cell deaths. For example, haploinsufficiency of beclin 1 leads to a tumour predisposition phenotype, indicating that autophagy is tumour suppressive³². More direct evidence has been provided by cells whose apoptotic machinery has been inhibited: when mouse embryonic fibroblasts (MEFs) that are null for both *Bax* and *Bak* are treated with the apoptosis inducers staurosporine or etoposide they undergo a form of cell death that is associated with autophagosomes, is dependent on *Atg5* and beclin 1 and is inhibited by the autophagy/class III phosphatidylinositol-3 kinase inhibitor 3-methyladenine³³.

Furthermore, Lenardo's group found that caspase inhibition by zVAD-fmk (a general caspase inhibitor) in L929 cells results in

autophagy-dependent PCD³⁴, which has been proposed to be mediated by the selective degradation of catalase³⁵. On the one hand, this may alert us to the possibility that anti-apoptotic therapies carry the potential risk of inducing non-apoptotic PCD; on the other hand, it may argue that therapeutics directed at multiple cell-death pathways will be required for optimal efficacy in diseases that involve PCD. Because cell death associated with some neurodegenerative diseases (or disease-associated mutants, such as α -synuclein) is associated with an autophagic morphology, these implications are potentially important for the development of effective therapies to prevent or ameliorate neurodegenerative disorders³⁶.

It is clear that *in vivo* testing to establish whether the autophagic process is required for any form(s) of PCD is needed, and the required genetic and pharmacological tools are increasingly available. Many questions in this area remain unanswered. If autophagy is a cellular-protective programme that — like the UPR — at some point activates PCD, what is the signal that initiates PCD? How important is the role of autophagic PCD in neurodegeneration? Does autophagic PCD occur *in vivo* in the absence of apoptosis inhibition? Are there 'executioners' analogous to caspases in autophagic PCD? Given the common finding of protein aggregates in neurodegenerative diseases, is a defect in autophagy a common underlying problem in these diseases? If so, does this contribute to the triggering of cell death in neurodegenerative diseases?

Alternative cell-death programmes

In comparison with apoptosis, little is known about autophagic PCD, and even less is known about other non-apoptotic forms of PCD. In fact,

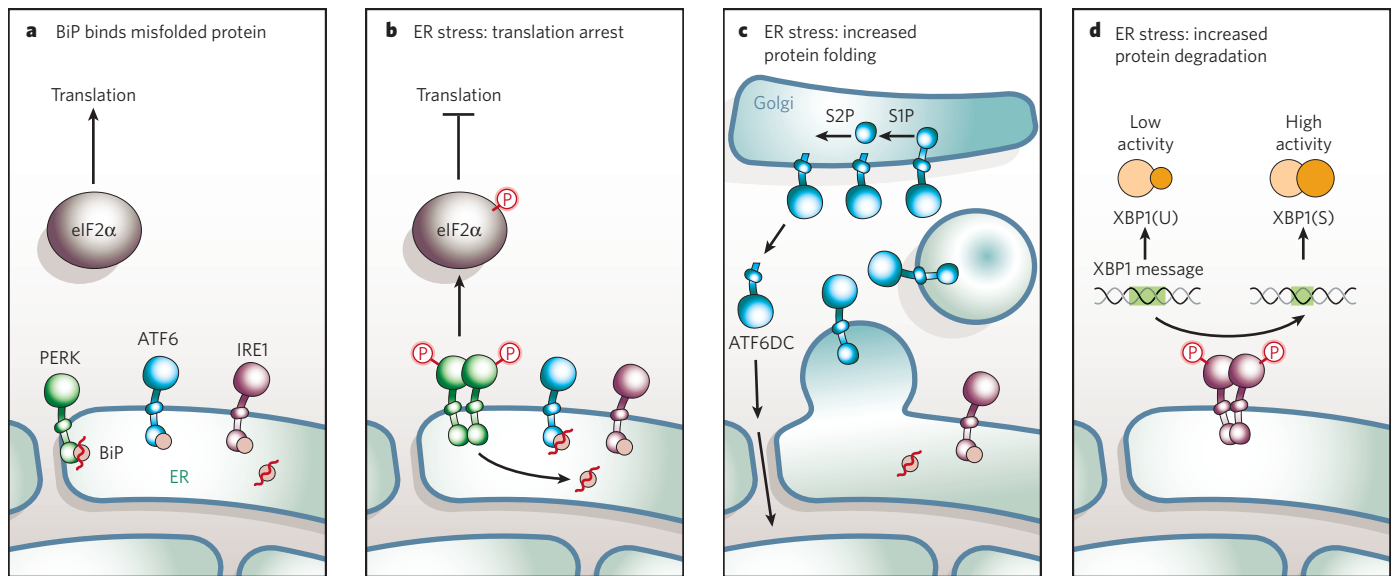


Figure 1 | Misfolded proteins and endoplasmic-reticulum stress. **a**, In addition to its role as an ER chaperone protein, BiP also regulates the activation of the three proximal ER-stress transducers — IRE1, PERK and ATF6. All three transducers contain an amino-terminal luminal domain that interacts with BiP. Under normal conditions, BiP serves as a negative regulator of IRE1, PERK and ATF6 activation. When misfolded proteins accumulate and result in ER stress, BiP binds to these proteins. BiP is thus released from the transducers, which are consequently activated. The activation of all three proximal sensors results in reduction in the amount of new protein translocated into the ER lumen, increased degradation of ER-localized proteins and increased protein-folding capacity of the ER. **b**, BiP release from PERK correlates with oligomerization, trans-autophosphorylation and activation of downstream signalling by PERK. PERK-dependent phosphorylation of eIF2 α on Ser 51 leads to attenuation of protein translation and thereby reduces the workload of the ER. **c**, BiP release from ATF6 permits transport of ATF6 to the Golgi compartment for regulated intramembrane proteolysis by site-1 protease (S1P) and site-2 protease (S2P) to generate a 50-kDa cytosolic b-ZIP-containing fragment (ATF6DC). ATF6DC translocates to the nucleus to activate ER-stress inducible target genes that code for proteins involved in protein folding. This BiP-regulated activation provides a direct mechanism to sense the folding capacity of the ER. **d**, Accumulation of misfolded proteins in the ER causes BiP to release IRE1. IRE1 is a type I transmembrane protein that contains both a serine/threonine kinase domain and an endoribonuclease domain; the latter processes an intron from X-box protein 1 (XBP1) mRNA, rendering it competent for translation to produce the 41-kDa XBP1(S) protein. Unspliced XBP1 mRNA encodes a basic leucine-zipper protein, XBP1(U), that lacks transactivation activity and is more labile than XBP1(S). XBP1(S) binds to the promoters of several genes involved in retrograde transport of misfolded proteins from the ER to the cytosol and in ER-induced protein degradation.

so far, other forms of PCD have not been generally accepted by the scientific community. However, as has previously been noted³⁷, such forms of cell death have been observed repeatedly, although their genetics and biochemical pathways are poorly understood. Type III PCD⁵, or cytoplasmic cell death, is a 'necrosis-like' form of PCD that includes swelling of the ER and mitochondria, and lacks typical apoptotic features such as apoptotic bodies and nuclear fragmentation. It has recently been noted that the hyperactivation of the tyrosine-kinase receptor insulin-like growth factor I receptor (IGFR) induces a non-apoptotic form of cell death known as paraptosis⁹. This was shown to be programmatic — in that it required transcription and translation — and was found to be morphologically indistinguishable from type III PCD. Neither Bcl-2 nor caspase inhibitors block this form of PCD, nor are caspases activated, but inhibitors of extracellular signal-regulated kinase 2 (ERK2) — but not ERK1 — were found to inhibit paraptosis³⁸, as was AIP-1 (a PCD-interacting protein also known as ALIX).

The idea that PCD might be induced by hyperactivation of a trophic factor receptor — trophotoxicity — is compatible with earlier observations that some trophic factors may increase neuronal cell death, for example that induced by excitotoxicity³⁹. Such an effect might be protective against neoplasia, in that it may eliminate cells that would otherwise undergo autocrine-loop-stimulated oncogenesis. The resulting programme would necessarily be non-apoptotic, because trophic factors inactivate apoptotic signalling.

Aponecrosis is a term applied to a combination of apoptosis and necrosis⁴⁰. Many cytotoxins induce PCD at low concentrations but at higher concentrations induce necrotic cell death, presumably because the cell's homeostatic processes are overwhelmed before the cell-death programmes can be completed. In fact, this is the most common pattern seen with cellular toxins, from hydrogen peroxide and other oxidants

to mitochondrial toxins such as antimycin A⁴⁰. Nicotera, Lipton and their colleagues showed that glutamate-induced neuronal-cell death could proceed through apoptosis or necrosis, depending on mitochondrial membrane potential and cellular energy state⁴¹. However, in most cases, the necrotic morphology associated with aponecrosis has not been proved not to be programmed. So, it is still not clear whether aponecrosis represents a combination of apoptosis and a non-apoptotic form of PCD, or whether it represents a combination of apoptosis and non-programmatic cell death.

Forms of cell death have been described that do not fit the criteria for any of the three types of developmental cell death (Table 1). For example, a non-apoptotic, caspase-independent form that does not resemble type II or type III developmental PCD has been described by Driscoll and her colleagues⁴² in *C. elegans* expressing mutant channel proteins such as MEC-4(d). A uniform, necrosis-like cell death — characterized morphologically by membranous whorls not seen in other types of cell death — is triggered by calcium entry, mediated by specific calpains and cathepsins, and inhibited by calreticulin.

A fifth apparent form of PCD has been described by the Dawsons and their colleagues, who showed that a non-apoptotic form of cell death depends on the activation of poly-(ADP-ribose) polymerase (PARP) and the consequent translocation of apoptosis-inducing factor (AIF) from the mitochondria to the nucleus⁴³. AIF is a flavoprotein described by Kroemer and colleagues⁴⁴ that is involved with DNA fragmentation, along with endonuclease G and DNA-fragmentation factor. This form of PCD was shown to be activated by agents that induce DNA damage, and shows a morphology and biochemistry that is, as far as we know, distinct from PCD types I–III.

It is likely that, as additional data are gathered from other cell-death paradigms, novel biochemical pathways of PCD will be characterized.

Box 2 | Endoplasmic-reticulum stress and cell death

ER stress is coupled to specific independent death pathways, as well as demonstrating involvement with the intrinsic and extrinsic apoptotic pathways. Studies from various laboratories have disclosed the roles of several ER-stress-induced cell-death modulators and effectors through the use of biochemical, pharmacological and genetic tools. These ER-stress-induced cell-death modulators include various members of the Bcl-2 family (Bcl-2, Bcl-X_L, BAX, BAK, BI-1 and BIK), BAP31 and p53-dependent gene products such as NOXA and PUMA.

BAP31, an ER-membrane protein, binds Bcl-2 (or Bcl-X_L) and a caspase-8-containing pro-apoptotic complex⁷⁷. When BAP31 is cleaved, a pro-apoptotic p20 fragment is derived, which, among other effects, induces mitochondrial fission, enhancing cytochrome c release⁷⁸. Conversely, BAR, which is expressed primarily in neurons of the central nervous system, also bridges Bcl-2 and caspase-8 but functions as an anti-apoptotic protein⁷⁹.

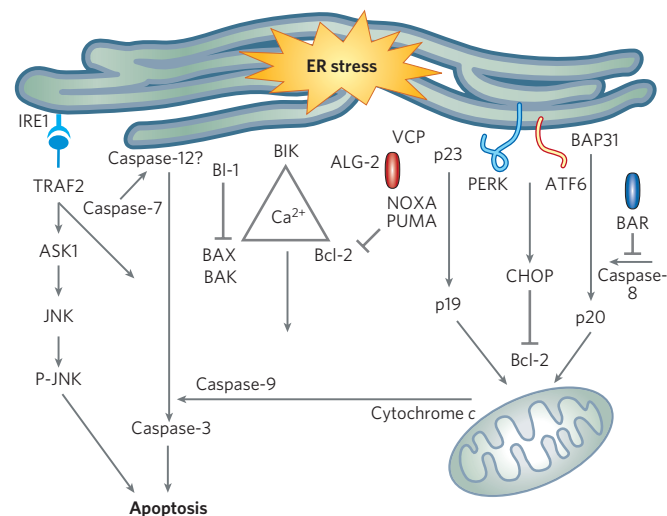
CHOP (also known as GADD153), a transcription factor induced during ER stress and activated by p38 mitogen-activated protein kinase, may also function as an ER-stress-induced cell-death modulator. It has been proposed to promote ER-stress-induced cell death by downregulating Bcl-2 expression.

Recent data have also implicated the calcium-binding protein apoptosis-linked gene-2 (ALG-2), and valosin-containing protein (VCP) as mediators of ER-stress-induced PCD. p23 interacts with PUMA and inhibits apoptosis, an effect that is reversed after cleavage. An ALG-2-interacting protein known as ALIX (or AIP-1) links motor-neuron-cell death during development as well as neuronal death in a HD model to the endolysosomal system⁸⁰.

In addition to the two main pathways that initiate the caspase cascade — namely, the extrinsic and intrinsic pathways — some studies point

to a caspase-12-mediated apoptotic pathway that involves the ER and the UPR (at least in murine cells), although this is a controversial area of research. Caspase-12 may be activated by calpain or caspase-7, or by ER-stress-activated IRE1, which may recruit caspase-12 through tumour-necrosis-factor receptor-associated factor 2 protein (TRAF2).

Oligomerized IRE1 also binds TRAF2, signalling downstream kinases (ASK1 and JNK) that, in turn, cause activation of pro-apoptotic proteins and cell death.



For example, there is an extensive literature on the morphological criteria for another potential form of PCD — oncosis — but the biochemical underpinnings of oncosis have not yet been described. Oncosis refers to a specific morphology of cell death — cellular swelling — that is typically induced by ischaemia and is thought to be mediated by the failure of plasma-membrane ionic pumps. One potential mediator of oncosis is a calpain-family protease (possibly a mitochondrial calpain⁴⁵).

Triggering cell death in neurodegeneration

As noted above, a number of different and potentially interrelated insults may occur as part of the neurodegenerative process. The role of the cell-death machinery in neurodegenerative diseases is controversial,

especially because synaptic loss and electrophysiological abnormalities typically precede cell loss in these disease states (see page 768). However, recent studies suggest a critical role for cell-death mediators in neurodegenerative diseases, even before the reduction in neuronal number. For example, in a transgenic mouse model of AD that features senile plaques, synapse and memory loss, but little or no neuronal loss, mutation of the caspase cleavage site at Asp 664 in the amyloid precursor protein (APP) completely suppressed synapse loss, dentate gyrus atrophy, astrogliosis and memory loss, even though senile plaque number and amyloid- β concentrations were unaffected⁴⁶. In an analogous study with an HD-transgenic mouse model, mutation of the caspase-6 site (but not the caspase-3 sites) in polyglutamine-expanded huntingtin prevented both the neurodegeneration and motor abnormalities characteristic of Huntington's disease⁴⁷.

More direct evidence for caspase activation in neurodegeneration (Fig. 2) comes from the employment of antibodies directed against neo-epitopes derived by caspase cleavage^{46,48} and from the inhibition of neurodegeneration by caspase inhibitors^{49,50}. However, some neurodegenerative models and diseases clearly demonstrate non-apoptotic forms of PCD⁷. Determining which PCD pathways are triggered in each neurodegenerative disease, which pathway accounts for each fraction of cell death, the mechanism(s) by which each pathway is triggered, and the interactions between the various pathways should shed new light on the degenerative process and its potential treatment or prevention. Perhaps even more important will be dissecting the pathways mediating subapoptotic events such as synaptosis and Wallerian degeneration — likely to be important features in the neurodegenerative process — and understanding how the interplay between these various processes results in the neurodegenerative phenotype.

The ability to initiate the neurodegenerative process, with widely varying insults — from misfolded proteins to reactive oxygen species to caspase recruitment complexes, as well as other mechanisms — and yet produce a relatively small number of syndromes indicates the existence of a death network that can be entered from many different sites but once triggered follows similar interdependent biochemical pathways with little dependence on the point of entry. This idea is compatible with

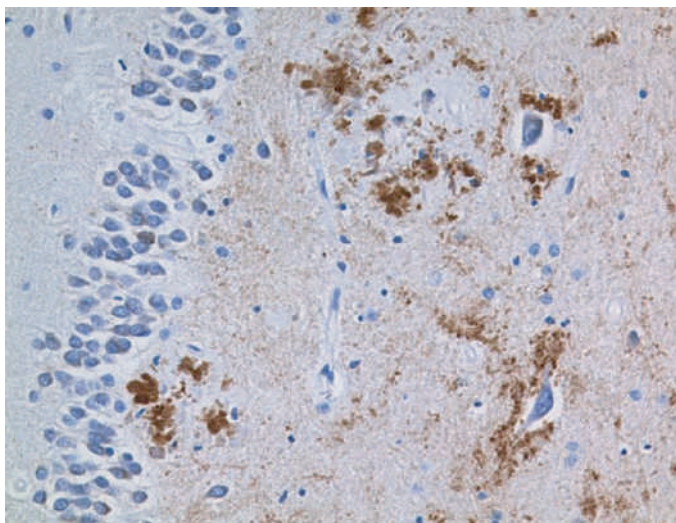


Figure 2 | Caspase cleavage in Alzheimer's disease. An antibody specific for the neoepitope generated by cleavage of APP at the caspase-cleavage site, Asp 664, demonstrates reactivity (brown) in the hippocampus of a patient with Alzheimer's disease.

the findings that therapeutics aimed at different pathways (for example, caspase activation, mitochondrial release of cytochrome *c*, metal binding and reactive-oxygen-species scavenging) all have partly salutary effects. However, it also suggests that a complete halt of the neurodegenerative process may require therapeutics that address all the interacting pathways of the network.

Targeting programmed cell death

Neuronal loss is a relatively late event in neurodegenerative diseases, following neuronal dysfunction, synapse loss and, often, somal atrophy. But targeting PCD has been successful in at least some model systems of neurodegeneration, and it is possible that targeting multiple PCD pathways will be even more effective. Caspase inhibition *in vivo* retarded the degeneration in transgenic mouse models of both ALS and HD, despite the fact that the morphological description of neuronal cell death in these diseases is not compatible with apoptosis^{49,50}. Bcl-2 expression in a transgenic model of ALS delayed symptom onset and increased lifespan, but did not alter the disease duration⁵¹.

Minocycline, a second-generation tetracycline that inhibits mitochondrial cytochrome *c* release, effected neuroprotection in mouse models of HD, PD and ALS⁵². Minocycline is orally bioavailable, penetrates the blood–brain barrier and has been proved safe for use in humans. It is currently being evaluated in clinical trials in patients with HD and ALS⁵².

Rasagiline, which has been approved for the treatment of PD, has also been proposed to target apoptosis, but because it is a potent, selective, irreversible inhibitor of monoamine oxidase type B, its therapeutic effect on PD may have nothing to do with effects on apoptosis.

Trophic factors have multiple effects, including the inhibition of apoptosis, the stimulation of neural precursors and the stimulation of neurite outgrowth. The literature is rife with examples of trophic-factor therapy for various neurological conditions, and although many have been unsuccessful, the delivery of the correct factor(s) to the right target in the correct concentration for a particular disease still holds great promise (although hyperactivation of at least some trophic-factor receptors may induce PCD). A recent study of fibroblast growth factor 2 (FGF2) in a mouse model of HD prolonged survival, improved motor performance and reduced polyglutamine aggregates⁵³. A number of approaches have been taken to deliver nerve growth factor (NGF) to patients with AD, the most recent being by means of genetically engineered fibroblasts in a phase I trial⁵⁴. In this study, improvements in cognition and positron-emission tomography (used to monitor fluorodeoxyglucose uptake) were documented. Glial-derived neurotrophic factor (GDNF) has shown promise in the treatment of PD⁵⁵.

HD may represent one of the best possibilities for trophic-factor therapy: presymptomatic diagnosis is readily available, and although a number of different trophic factors have shown promise in animal models (for example, FGF2, NGF and ciliary neurotrophic factor⁵⁶), brain-derived neurotrophic factor (BDNF) is both reduced in the disease and therapeutic in models. Furthermore, cysteamine has been shown to increase BDNF levels in brain⁵⁷.

A lifeline

Exciting recent evidence suggests that pathological processes may stimulate neurogenesis in the brain and may redirect the migration of nascent neurons towards the site of pathology. And, under certain circumstances, stimulating neurogenesis can improve the performance and survival of mice with neurodegenerative disease. These and similar results raise the question of whether stem-cell exhaustion or senescence after prolonged stimulation might have a role in the long-term course of neurodegenerative disease. Furthermore, although the factors linking neurodegeneration to neural-stem-cell proliferation and inhibition of PCD are largely unknown, the potential for therapy using these putative factors is likely to be significant. Reconciliation of cell-death pathways with neurodegenerative and regenerative mechanisms should offer an improved understanding of disease and open avenues for therapeutic intervention. ■

1. Studnicka, F. K. In *Lehrbuch der vergleichende mikroskopischen Anatomie der Wirbeltiere* (ed. Oppel, A.) 1–256 (Fischer, Germany, 1905).
2. Levi-Montalcini, R. The nerve growth factor: its mode of action on sensory and sympathetic nerve cells. *Harvey Lect.* **60**, 217–259 (1966).
3. Lockshin, R. A. & Williams, C. M. Programmed cell death. II. Endocrine potentiation of the breakdown of the intersegmental muscles of silkworms. *J. Insect Physiol.* **10**, 643–649 (1964).
4. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
5. Clarke, P. G. Developmental cell death: morphological diversity and multiple mechanisms. *Anat. Embryol.* **181**, 195–213 (1990).
6. Cunningham, T. J. Naturally occurring neuron death and its regulation by developing neural pathways. *Int. Rev. Cytol.* **74**, 163–186 (1982).
7. Dal Canto, M. C. & Gurney, M. E. Development of central nervous system pathology in a murine transgenic model of human amyotrophic lateral sclerosis. *Am. J. Pathol.* **145**, 1271–1279 (1994).
8. Schweichel, J. U. & Merker, H. J. The morphology of various types of cell death in prenatal tissues. *Teratology* **7**, 253–266 (1973).
9. Sperandio, S., de Belle, I. & Bredesen, D. E. An alternative, non-apoptotic form of programmed cell death. *Proc. Natl Acad. Sci. USA* **97**, 14376–14381 (2000).
10. Oppenheim, R. W. Naturally occurring cell death during neural development. *Trends Neurosci.* **17**, 487–493 (1985).
11. Fadok, V. A. et al. Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages. *J. Immunol.* **148**, 2207–2216 (1992).
12. Thornberry, N. A. & Lazebnik, Y. Caspases: enemies within. *Science* **281**, 1312–1316 (1998).
13. Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. The *C. elegans* cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1 β -converting enzyme. *Cell* **75**, 641–652 (1993).
14. Salvesen, G. S. & Dixit, V. M. Caspases: intracellular signaling by proteolysis. *Cell* **91**, 443–446 (1997).
15. Morishima, N., Nakanishi, K., Takenouchi, H., Shibata, T. & Yasuhiko, Y. An endoplasmic reticulum stress-specific caspase cascade in apoptosis. Cytochrome *c*-independent activation of caspase-9 by caspase-12. *J. Biol. Chem.* **277**, 34287–34294 (2002).
16. Rao, R. V. et al. Coupling endoplasmic reticulum stress to the cell death program. An Apaf-1-independent intrinsic pathway. *J. Biol. Chem.* **277**, 21836–21842 (2002).
17. Yuan, J. & Yankner, B. A. Caspase activity sows the seeds of neuronal death. *Nature Cell Biol.* **1**, E44–E45 (1999).
18. Green, D. R. & Kroemer, G. Pharmacological manipulation of cell death: clinical applications in sight? *J. Clin. Invest.* **115**, 2610–2617 (2005).
19. Fuentes-Prior, P. & Salvesen, G. S. The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem. J.* **384**, 201–232 (2004).
20. Kopito, R. R. & Ron, D. Conformational disease. *Nature Cell Biol.* **2**, E207–E209 (2000).
21. Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *Science* **296**, 1991–1995 (2002).
22. Sitia, R. & Braakman, I. Quality control in the endoplasmic reticulum protein factory. *Nature* **426**, 891–894 (2003).
23. Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biology thinks about neurodegenerative diseases. *Neuron* **29**, 15–32 (2001).
24. Rao, R. V. & Bredesen, D. E. Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr. Opin. Cell Biol.* **16**, 653–662 (2004).
25. Scorrano, L. et al. BAX and BAK regulation of endoplasmic reticulum Ca²⁺: a control point for apoptosis. *Science* **300**, 135–139 (2003).
26. Ruiz-Vela, A., Opferman, J. T., Cheng, E. H. & Korsmeyer, S. J. Proapoptotic BAX and BAK control multiple initiator caspases. *EMBO Rep.* **6**, 379–385 (2005).
27. Chae, H. J. et al. BI-1 regulates an apoptosis pathway linked to endoplasmic reticulum stress. *Mol. Cell* **15**, 355–366 (2004).
28. Li, J., Lee, B. & Lee, A. S. Endoplasmic reticulum stress-induced apoptosis: multiple pathways and activation of p53-up-regulated modulator of apoptosis (PUMA) and NOXA by p53. *J. Biol. Chem.* **281**, 7260–7270 (2006).
29. Hegde, R. S. et al. A transmembrane form of the prion protein in neurodegenerative disease. *Science* **279**, 827–834 (1998).
30. Levine, B. & Yuan, J. Autophagy in cell death: an innocent convict? *J. Clin. Invest.* **115**, 2679–2688 (2005).
31. Komatsu, M. et al. Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
32. Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100**, 15077–15082 (2003).
33. Shimizu, S. et al. Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes. *Nature Cell Biol.* **6**, 1221–1228 (2004).
34. Yu, L. et al. Regulation of an ATG7-beclin 1 program of autophagic cell death by caspase-8. *Science* **304**, 1500–1502 (2004).
35. Yu, L. et al. Autophagic programmed cell death by selective catalase degradation. *Proc. Natl Acad. Sci. USA* **103**, 4952–4957 (2006).
36. Gomez-Santos, C. et al. Dopamine induces autophagic cell death and α -synuclein increase in human neuroblastoma SH-SY5Y cells. *J. Neurosci. Res.* **73**, 341–350 (2003).
37. Hengartner, M. O. The biochemistry of apoptosis. *Nature* **407**, 770–776 (2000).
38. Sperandio, S. et al. Paraptosis: mediation by MAP kinases and inhibition by AIP-1/Alix. *Cell Death Differ.* **11**, 1066–1075 (2004).
39. Koh, J. Y., Gwag, B. J., Lobner, D. & Choi, D. W. Potentiated necrosis of cultured cortical neurons by neurotrophins. *Science* **268**, 573–575 (1995).
40. Formigli, L. et al. Aponecrosis: morphological and biochemical exploration of a syncretic process of cell death sharing apoptosis and necrosis. *J. Cell Physiol.* **182**, 41–49 (2000).
41. Ankarcrona, M. et al. Glutamate-induced neuronal death: a succession of necrosis or apoptosis depending on mitochondrial function. *Neuron* **15**, 961–973 (1995).
42. Syntichaki, P., Xu, K., Driscoll, M. & Tavernarakis, N. Specific aspartyl and calpain proteases are required for neurodegeneration in *C. elegans*. *Nature* **419**, 939–944 (2002).

43. Yu, S. W. *et al.* Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor. *Science* **297**, 259–263 (2002).
44. Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**, 441–446 (1999).
45. Liu, X., Van Vleet, T. & Schnellmann, R. G. The role of calpain in oncotic cell death. *Annu. Rev. Pharmacol. Toxicol.* **44**, 349–370 (2004).
46. Galvan, V. *et al.* Reversal of Alzheimer's-like pathology and behavior in human APP transgenic mice by mutation of Asp664. *Proc. Natl Acad. Sci. USA* **103**, 7130–7135 (2006).
47. Graham, R. K. *et al.* Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell* **125**, 1179–1191 (2006).
48. Yang, F. *et al.* Antibody to caspase-cleaved actin detects apoptosis in differentiated neuroblastoma and plaque-associated neurons and microglia in Alzheimer's disease. *Am. J. Pathol.* **152**, 379–389 (1998).
49. Friedlander, R. M., Brown, R. H., Gagliardini, V., Wang, J. & Yuan, J. Inhibition of ICE slows ALS in mice. *Nature* **388**, 31 (1997).
50. Ona, V. O. *et al.* Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* **399**, 263–267 (1999).
51. Kostic, V., Jackson-Lewis, V., de Bilbao, F., Dubois-Dauphin, M. & Przedborski, S. Bcl-2: prolonging life in a transgenic mouse model of familial amyotrophic lateral sclerosis. *Science* **277**, 559–562 (1997).
52. Friedlander, R. M. Apoptosis and caspases in neurodegenerative diseases. *N. Engl. J. Med.* **348**, 1365–1375 (2003).
53. Jin, K. *et al.* FGF-2 promotes neurogenesis and neuroprotection and prolongs survival in a transgenic mouse model of Huntington's disease. *Proc. Natl Acad. Sci. USA* **102**, 18189–18194 (2005).
54. Tuszynski, M. H. *et al.* A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease. *Nature Med.* **11**, 551–555 (2005).
55. Lang, A. E. *et al.* Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease. *Ann. Neurol.* **59**, 459–466 (2006).
56. Kordower, J. H., Isacson, O. & Emerich, D. F. Cellular delivery of trophic factors for the treatment of Huntington's disease: is neuroprotection possible? *Exp. Neurol.* **159**, 4–20 (1999).
57. Borrell-Pages, M. *et al.* Cystamine and cysteamine increase brain levels of BDNF in Huntington disease via HSP1b and transglutaminase. *J. Clin. Invest.* **116**, 1410–1424 (2006).
58. Kerr, J. F. R. & Harmon, B. V. in *Apoptosis: The Molecular Basis of Cell Death* (eds Tomei, L. D. & Cope, F. O.) 321 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1991).
59. Bursch, W. *et al.* Autophagic and apoptotic types of programmed cell death exhibit different fates of cytoskeletal filaments. *J. Cell Sci.* **113**, 1189–1198 (2000).
60. Hall, I. H., Elkins, A. L., Karthikeyan, S. & Spielvogel, B. F. The cytotoxicity of 1-(phenylmethyl)-4,7,10-tris-[(4-methylphenyl) sulfonyl]-1,4,7,10-tetraazacyclododecane in human Tmol3 T leukemic cells. *Anticancer Res.* **17**, 1195–1198 (1997).
61. Susin, S. A. *et al.* Two distinct pathways leading to nuclear apoptosis. *J. Exp. Med.* **192**, 571–580 (2000).
62. Ohno, M. *et al.* 'Apoptotic' myocytes in infarct area in rabbit hearts may be oncotic myocytes with DNA fragmentation: analysis by immunogold electron microscopy combined with *in situ* nick end-labeling. *Circulation* **98**, 1422–1430 (1998).
63. Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
64. Kuwana, T. *et al.* Bid, Bax, and lipids cooperate to form supramolecular openings in the outer mitochondrial membrane. *Cell* **111**, 331–342 (2002).
65. Guo, B. *et al.* Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature* **423**, 456–461 (2003).
66. Schuler, M., Bossy-Wetzel, E., Goldstein, J. C., Fitzgerald, P. & Green, D. R. p53 induces apoptosis by caspase activation through mitochondrial cytochrome c release. *J. Biol. Chem.* **275**, 7337–7342 (2000).
67. Lin, B. *et al.* Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor Nur77/TR3. *Cell* **116**, 527–540 (2004).
68. Deveraux, Q. L., Takahashi, R., Salvesen, G. S. & Reed, J. C. X-linked IAP is a direct inhibitor of cell-death proteases. *Nature* **388**, 300–304 (1997).
69. Holley, C. L., Olson, M. R., Colon-Ramos, D. A. & Kornbluth, S. Reaper eliminates IAP proteins through stimulated IAP degradation and generalized translational inhibition. *Nature Cell Biol.* **4**, 439–444 (2002).
70. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
71. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
72. Martins, L. M. *et al.* The serine protease Omi/HtrA2 regulates apoptosis by binding XIAP through a reaper-like motif. *J. Biol. Chem.* **277**, 439–444 (2002).
73. Bossy-Wetzel, E., Barsom, M. J., Godzik, A., Schwarzenbacher, R. & Lipton, S. A. Mitochondrial fission in apoptosis, neurodegeneration and aging. *Curr. Opin. Cell Biol.* **15**, 706–716 (2003).
74. Lee, Y. J., Jeong, S. Y., Karbowski, M., Smith, C. L. & Youle, R. J. Roles of the mammalian mitochondrial fission and fusion mediators Fis1, Drp1, and Opa1 in apoptosis. *Mol. Biol. Cell.* **15**, 5001–5011 (2004).
75. Frezza, C. *et al.* OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell* **126**, 177–189 (2006).
76. Cipolat, S. *et al.* Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**, 163–175 (2006).
77. Ng, F. W. *et al.* p28 Bap31, a Bcl-2/Bcl-X_L- and procaspase-8-associated protein in the endoplasmic reticulum. *J. Cell Biol.* **139**, 327–338 (1997).
78. Breckenridge, D. G., Stojanovic, M., Marcellus, R. C. & Shore, G. C. Caspase cleavage product of BAP31 induces mitochondrial fission through endoplasmic reticulum calcium signals, enhancing cytochrome c release to the cytosol. *J. Cell Biol.* **160**, 1115–1127 (2003).
79. Roth, W. *et al.* Bifunctional apoptosis inhibitor (BAR) protects neurons from diverse cell death pathways. *Cell Death Differ.* **10**, 1178–1187 (2003).
80. Mahul-Mellier, A. L., Hemming, F. J., Blot, B., Fraboulet, S. & Sadoul, R. Alix, making a link between apoptosis-linked gene-2, the endosomal sorting complexes required for transport, and neuronal death *in vivo*. *J. Neurosci.* **26**, 542–549 (2006).

Acknowledgements We apologize to all colleagues whose papers we were unable to cite owing to space limitations. We thank S. Rabizadeh, A. Kurakin, V. Galvan, D. Madden, L. Egger, J. Fombonne, T.-V. Nguyen, N. Rooke, K. Niazi, A. Swistowski, K. Poksay and S. Chen for critical reading of the manuscript, and M. Susag, L. Sheridan and R. Abulencia for manuscript preparation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to D.E.B. (dbredesen@buckinstitute.org).

Prions and their partners in crime

Byron Caughey¹ & Gerald S. Baron¹

Prions, the infectious agents of transmissible spongiform encephalopathies (TSEs), have defied full characterization for decades. The dogma has been that prions lack nucleic acids and are composed of a pathological, self-inducing form of the host's prion protein (PrP). Recent progress in propagating TSE infectivity in cell-free systems has effectively ruled out the involvement of foreign nucleic acids. However, host-derived nucleic acids or other non-PrP molecules seem to be crucial. Interactions between TSE-associated PrP and its normal counterpart are also pathologically important, so the physiological functions of normal PrP and how they might be corrupted by TSE infections have been the subject of recent research.

The transmissible spongiform encephalopathies (TSEs) first attracted widespread scientific attention almost half a century ago. The infectious agents responsible for TSEs were unusually hard to eliminate, and were more resistant to heat and radiation than any known pathogen. This led to heretical proposals that they might lack nucleic-acid genomes. The prototypical TSE disease was scrapie, a widespread infection of sheep that was known to persist in pastures for years after the removal of infected animals. Similarities between scrapie and human kuru, which was spread by cannibalistic feasts, contributed to a lurid fascination with these devastating and fatal neurodegenerative diseases. Since then, the impact of TSEs has been greatly amplified by the recognition of more common human forms such as Creutzfeldt–Jakob disease (CJD), and epidemics of bovine spongiform encephalopathy (BSE) and cervid chronic wasting disease (CWD).

TSEs (or prion diseases) are now defined by the transmission and involvement in neuropathogenesis of abnormal forms of prion protein (PrP) (reviewed in ref. 1). The normal form of PrP, PrP^C, is a glycoprotein that is usually monomeric in structure, protease-sensitive, and linked to cellular membranes through a glycosphosphatidylinositol (GPI) anchor (Box 1). Pathological forms of PrP vary and are poorly defined structurally, but are most commonly multimeric (aggregated) and more protease-resistant than PrP^C. They have been given various names that emphasize associations with disease (for example, PrP^{TSE} and PrP^d), infectivity (PrP^{Sc}), toxicity (PrP^{tox} and PrP^L) and relative protease-resistance (PrPres). Each of these terms has limitations. Here, we primarily use the term PrPres to refer to partly proteinase-K-resistant, TSE-associated forms of PrP, bearing in mind that most, but not all, types of TSE infectivity have been linked to such forms of PrP. The main difference between PrP^C and PrPres seems to be conformational, with PrPres having a higher β -sheet content and being multimeric. These characteristics, as well as its ability to assemble into amyloid fibrils, make PrPres similar to the abnormal protein aggregates that are associated with other protein misfolding diseases such as Alzheimer's disease, Parkinson's disease and Huntington's disease (see page 774). However, a key and intriguing difference is that PrPres preparations are infectious.

Although researchers have been able to associate PrPres with TSE infectivity (prions), the full molecular nature of the TSE agent remains unclear. This is largely due to difficulties in purifying the infective agent so that its essential components and structures can be determined. A long-suspected and widely assumed proposition is that mammalian prions are comprised solely of PrPres, which propagates itself by inducing the conversion of PrP^C to PrPres (ref. 1; Box 1). Nonetheless, definitive

proof of this 'protein-only' prion hypothesis has been lacking. Years ago, PrPres preparations were shown to convert PrP^C to PrPres with strain and species specificity¹, indicating that PrPres has at least some capacity to self-propagate. However, the yield of these simple conversion reactions was limited, and the newly made PrPres was not proved to be infectious (ref. 2; G. Raymond and B.C., unpublished observations).

Considerable progress has recently been made in the cell-free propagation of PrPres and TSE infectivity, making it highly unlikely that this process relies on replication of a virus or foreign nucleic acid. Nonetheless, the data so far fall short of proving that the agent is composed solely of PrPres. Indeed, as detailed below, it seems likely that other molecules are crucial for prion propagation, either as components or cofactors in the conformational conversion of PrP^C to PrPres. Furthermore, PrP^C may be a key mediator of the neurotoxicity of PrPres, highlighting the importance of investigating how the normal functions of PrP^C might be corrupted by TSE infection. In this review, we attempt to integrate these findings, with additional consideration given to common features of various anti-TSE compounds and how they might relate to the physiological ligands of PrP^C.

Unravelling TSE prion composition

The ultimate experiment to prove the nature of mammalian prions would be the synthesis of prions *in vitro* from defined constituents. Recently, three groups have come close to achieving this feat. One team showed that synthetic truncated PrP fibril preparations can induce TSE disease when inoculated into transgenic mice that vastly overexpress the same truncated PrP construct^{3,4}. Although these results are tantalizing, the recombinant PrP fibril preparations^{3,4} seem to be >10⁸ times less efficient at instigating TSE disease than *bona fide* mouse scrapie infectivity when inoculated into wild-type mice. Moreover, crucial work to demonstrate that the transgenic assay mice were not already making prions spontaneously, as has been shown with transgenic mice overexpressing a different mutant PrP transgene⁵, remains to be reported. To further complicate matters, Nazor *et al.*⁶ found that in similar transgenic mice inoculation of mutant-PrP aggregates from one transgenic mouse into others promotes the accumulation of pre-existing pathological forms of mutant PrP produced as a result of transgene overexpression. Accordingly, this process was more aptly characterized as acceleration of inherent disease rather than prion transmission — an important distinction in evaluating studies of *in vitro*-generated TSE prions.

Castilla *et al.* have demonstrated the cell-free propagation of PrPres and accompanying robust TSE infectivity, but many molecules besides

¹National Institute of Allergy and Infectious Disease, National Institutes of Health, Rocky Mountain Laboratories, 903 South 4th Street, Hamilton, Montana 59840, USA.

Box 1 | The cell biology of transmissible spongiform encephalopathies

As a GPI-anchored plasma-membrane glycoprotein (panel **a**), the PrP^C polypeptide is synthesized in the endoplasmic reticulum (ER), processed in the Golgi apparatus and then carried, in its mature form, to the cell surface (panel **b**). PrP^C is usually associated with detergent-resistant membrane subdomains known as rafts.

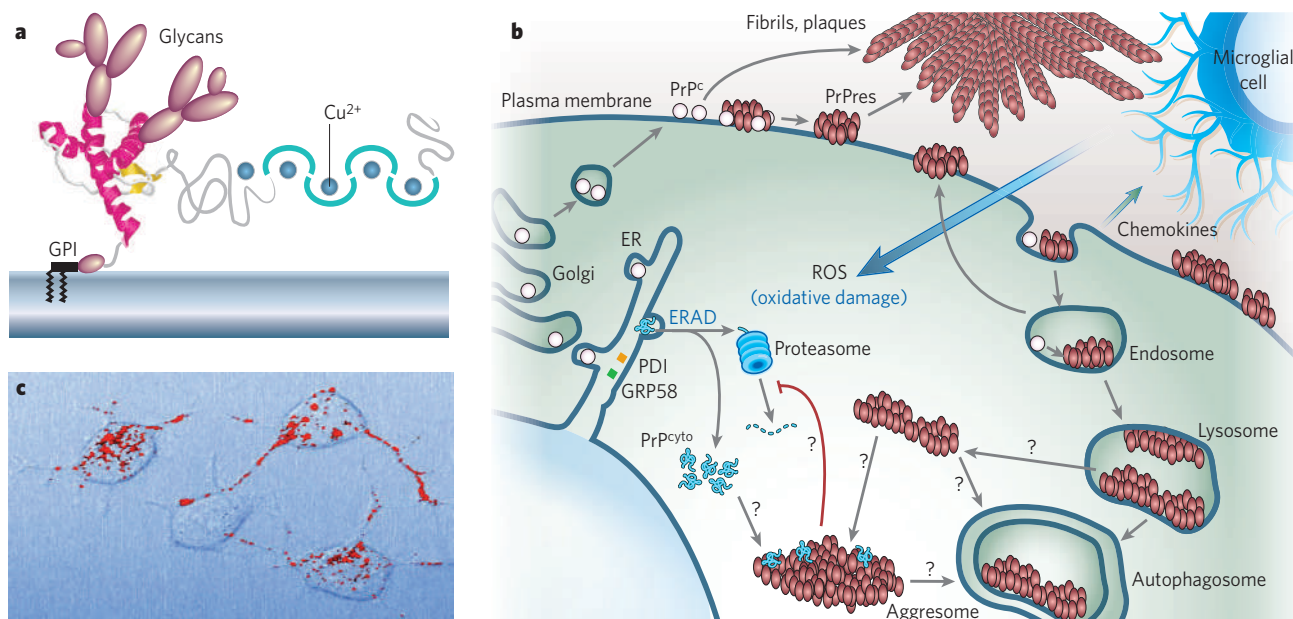
In cultured scrapie-infected cells, the generation of PrPres from PrP^C occurs after the arrival of PrP^C at the cell surface^{82,83}. This conversion is directly induced by pre-existing PrPres and apparent cofactors through an ill-defined templating mechanism that results in PrP^C joining the inducing PrPres oligomer or polymer. Important modulators of this conversion include sulphated GAG-containing proteoglycans such as HSPG (Box 2), raft membrane lipids⁸⁴ and LR/LRP¹⁶. The scrapie-associated conversion site for membrane-anchored wild-type PrP^C seems to be on the cell surface and/or in endosomes that are drawn in from the cell surface^{82,83,85}. However, PrP^C released from the cell due to lack of a GPI

anchor¹⁹ may be converted to extracellular deposits such as amyloid fibrils and plaques (Fig. 1). In familial TSE-associated mutant PrP^C molecules, spontaneous folding abnormalities can occur in the endoplasmic reticulum and/or the Golgi apparatus (for a review see ref. 30).

Misfolded PrP^C can be subject to the ER-associated degradation pathway (ERAD; panel **b**). Under conditions of proteasome inhibition, cytoplasmic forms of PrP aggregates associated with neurotoxicity (such as PrP^{Cyto} and aggresomes) have also been described^{22,52}, but the implications of these findings remain controversial. Furthermore, the mechanisms by which such aggregates are generated (for instance, the translocation of PrPres from the lumen of endocytic and lysosomal vesicles into cytosolic aggresomes, and the cause of proteasome inhibition) are unknown. The release of reactive oxygen species (ROS) from chemokine-activated microglial cells²⁶ could contribute to ER stress and/or the ERAD process by inactivation of ER-chaperones⁸⁶ (for example, protein disulphide isomerase (PDI)

and GRP58), one of which has been shown to protect against prion neurotoxicity²³. Excessive levels of misfolded proteins in the cytosol might impair proteasome function, either directly or after incorporation into aggresomes.

Once PrPres has been made, it can accumulate on the cell surface, in intracellular vesicles such as lysosomes or autophagosomes, or in extracellular deposits. PrPres and TSE infectivity can be released from cells in association with membrane-enclosed vesicles such as exosomes⁸⁷. Uninfected neuronal cells can incorporate PrPres from the extracellular milieu by means of a mechanism that involves HSPG and LRP/LR^{10,11,58,59}, distribute it along neuritic pathways, and initiate propagation of new PrPres⁷². Panel **c** shows a pseudo-coloured image of cultured neuronal cells to which fluorescently tagged scrapie PrPres (red) has been added. During the infection process, the cells have taken up PrPres particles into acidic vesicles and transported them along the neurites to points of contact with other cells⁷².



PrP were also present⁷. A cyclical conversion reaction (known as protein misfolding cyclic amplification; PMCA) was initiated by diluting scrapie-infected brain homogenate in normal brain homogenate (the latter provides the PrP^C substrate and other potential cofactors for PrPres amplification). The products were further diluted in normal brain homogenates for subsequent serial amplification cycles. That the infective agent of TSEs can replicate in a cell-free environment greatly strengthens the argument against prions being a viral agent requiring amplification of an exogenous pathogen-derived genome. However, the complexity of brain homogenates makes it difficult to establish which host molecule or molecules comprise the infectious prion 'unit'.

To begin to address this problem, Deleault *et al.* have shown that similar PrPres amplification can be supported using only PrPres, PrP^C purified from brain samples, and polyanionic accessory molecules such as nucleic acids and heparan sulphate proteoglycan (HSPG)⁸. This involvement of HSPG extends earlier observations that sulphated glycans can strongly influence (both positively and negatively) PrPres

formation *in vitro*⁹ (Box 2), co-localize with PrPres deposits *in vivo*, and serve as PrPres receptors on the surface of cells *in vitro*^{10,11}. Perhaps endogenous polyanions could associate with synthetic PrP fibrils after inoculation to bolster infectivity levels. However, it is not yet clear whether infectivity is generated in any of the polyanion-stimulated conversions of purified PrP^C.

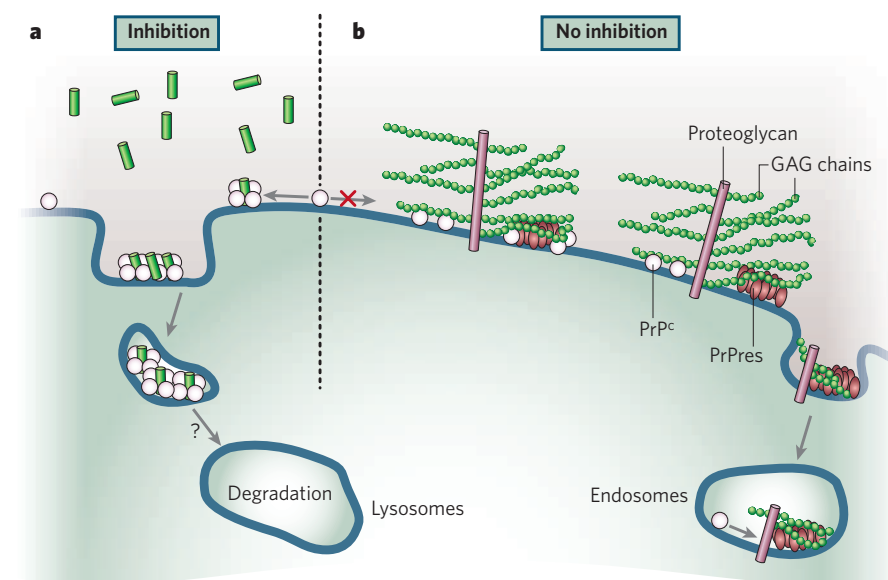
In summary, none of the recent reports provide conclusive evidence that the infective TSE agent is composed of protein alone. On the contrary, it seems just as plausible to argue that other host-derived molecules besides PrP, particularly polyanionic polymers such as sulphated glycosaminoglycans^{9,12,13} or host-derived nucleic acids^{8,14,15}, might be required for robust TSE infectivity. Certain other proteins¹⁶ also seem to be important, at least as cofactors in the conversion of normal PrP to pathological PrP in intact cells. Whether or not PrP molecules ultimately prove to be the only essential component for infectivity, the necessary conditions, cofactors, conformational changes and aggregation required for TSE-agent propagation *in vivo* remain to be determined.

Box 2 | Anti-TSE compounds

A variety of compounds have been identified that can inhibit PrPres formation in cell culture and promote survival in TSE-infected animals. These include cyclic tetrapyrroles^{88,89}, sulphated glycans (for example, GAGs), sulphonated dyes and phosphorothioated oligonucleotides (PS-ONs)⁵⁷ (reviewed in ref. 90). The most effective examples can more than triple the survival times of rodents inoculated intraperitoneally with high scrapie titres (for example, 10^3 – 10^4 times the lethal dose) and completely protect animals receiving lower titres. Under experimental conditions, a few compounds have been shown to have beneficial effects after infection of the central nervous system (reviewed in refs 88, 90, 91), but none are known to be effective against TSEs in clinical settings.

Many known anti-TSE compounds bind directly to PrP^C (and in some cases PrPres) and share a common structural feature in being (or having the capacity to assemble into) linear polyanionic chains⁹². This suggests that these compounds exert their inhibition by binding to PrP molecules at the same or overlapping sites. Indeed, competitive-binding studies have shown that sulphated glycans compete with Congo red, a sulphonated dye⁹³, and PS-ONs⁵⁷ for binding to PrP^C. Analyses of the sulphated-glycan-binding site on PrP^C have produced evidence for the involvement of residues in three different segments of the amino-acid sequence: the highly cationic amino-terminal residues, the octapeptide repeats and a more carboxy-terminal site containing residues 110–128. There is still some debate as to which residues are most important^{94–96}. The residues involved in binding different classes of anionic PrPres inhibitors may vary somewhat, depending on the size and specific nature of the particular inhibitor.

Regardless of the precise PrP-binding mechanism(s), one net effect of at least some



of these inhibitors^{57,65,95} is the clustering and internalization of PrP^C from the cell surface (see panel **a**). This results in the sequestration of PrP^C in a state and/or subcellular location that is incompatible with conversion to PrPres. Some inhibitors may also bind to PrPres in a way that interferes with its ability to convert PrP^C.

Because several structurally different classes of PrPres inhibitors share certain properties, PrP-binding sites, and tendencies to cause PrP aggregation and endocytosis, one wonders how these phenomena might relate to the normal function of PrP^C and the mechanism of conversion to PrPres. More specifically, it seems likely that these inhibitors bind to a site normally reserved for physiological ligands that influence the conversion to PrPres. Prime

candidates for this type of ligand are GAGs such as heparan sulphate that bind to PrP^C, associate with PrPres deposits *in vivo*, and support PrPres formation and internalization^{9–11,71} (panel **b**). Many of the PrPres inhibitors discussed above can be viewed as GAG analogues or mimics. These compounds also mimic nucleic acids, which, like GAGs, are often linear polyanionic polymers with repeating anionic and hydrophobic surfaces.

Finally, strong inhibition by various cyclic tetrapyrroles raises the possibility that PrP^C interacts normally with natural endogenous tetrapyrroles such as haemin. Crucial interactions with other cellular components such as laminin receptors⁶⁰ (Fig. 2) and membrane rafts might also be directly or indirectly affected by inhibitor binding.

Minimal and most infectious units

TSEs, like other neurodegenerative protein-misfolding diseases, generate various abnormal protein species, which range from non-fibrillar oligomers to large amyloid plaques. Questions abound as to which of these structures are the prime cause, or causes, of disease (see page 774). To better characterize the minimal infectious unit and the most infectious particles per unit protein, Silveira *et al.*¹⁷ developed new methods for gently disaggregating scrapie prion aggregates in detergent and separating them by size using flow field-flow fractionation. With respect to PrP content, marked overlapping peaks were seen for both infectivity and cell-free PrP conversion activity with 17–27 nm diameter non-fibrillar particles of molecular mass 300–600 kDa. These activities were substantially lower in larger fibrillar aggregates, and almost absent from oligomers of ≤ 5 PrP molecules. Thus, under the experimental conditions, non-fibrillar particles with masses equivalent to between 14 and 28 PrP molecules were the most efficient initiators of TSE disease. These observations are consistent with an emerging theme for many protein misfolding diseases, that small-to-intermediate abnormal protein oligomers are more pathogenic than their larger fibrillar amyloid counterparts¹⁸. Although PrP seemed to be the main protein of the minimal and most infectious TSE particles, it is not yet clear whether other molecules or non-protein constituents of PrP might also be essential to infectivity.

Post-translational modifications

One clear difference between the highly infectious PrPres generated *in vitro* by PMCA reactions⁷ and the purely synthetic recombinant PrP fibrils that have little or no infectivity³ is the presence of natural post-translational modifications such as N-linked glycans and GPI anchors (Box 1) on the PMCA PrPres. These non-protein attachments should influence the shapes, stabilities, interactions and activities of PrP molecules, and might be expected to be crucial in the formation of PrPres and infectivity. However, they do not seem to be essential, because both PrPres and TSE infectivity can be amplified in transgenic mice expressing only 'anchorless' PrP^C, which lacks a GPI anchor and is greatly underglycosylated¹⁹.

A role for GPI anchors

Although scrapie-infected anchorless-PrP transgenic mice propagate PrPres and infectivity, the lack of GPI-anchored PrP molecules changes the brain distribution of PrPres drastically and eliminates typical clinical TSE disease¹⁹. Instead of the predominantly non-amyloid deposits of PrPres seen in wild-type animals, the PrPres of anchorless PrP mice is restricted to large extracellular amyloid plaques (Fig. 1). Furthermore, the appearance and distribution of the neuropathological lesions that are associated with the anchorless PrPres amyloid differ from those seen in wild-type mice, with damage more prevalent in white matter than grey matter (B. Chesebro, personal communication). The lack of

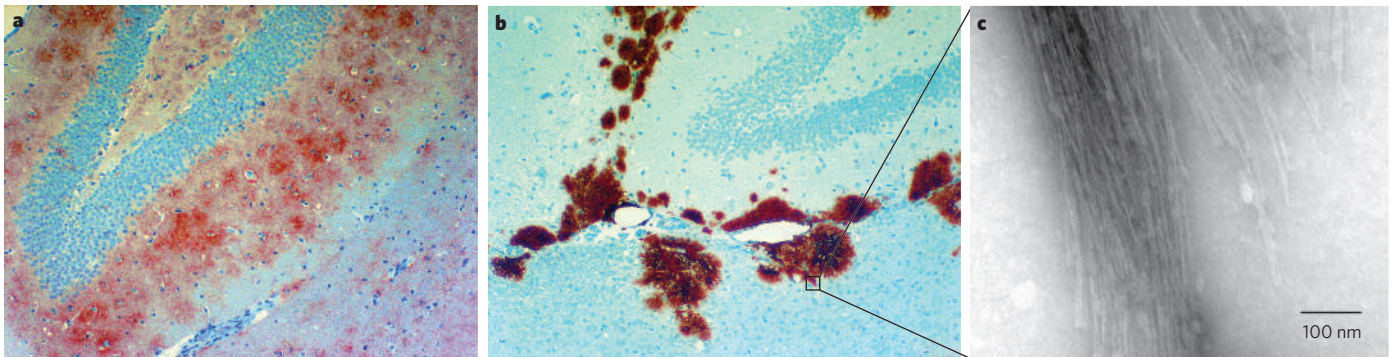


Figure 1 | Different patterns of PrPres deposition in brain tissue of mice expressing wild-type versus anchorless PrP^C. The sections shown are from the dentate gyrus region of the hippocampus. Two mouse strains were infected with the 22L strain of scrapie. **a**, PrPres distribution (red) in wild-type PrP^C mouse. **b**, PrPres distribution in anchorless PrP^C mouse¹⁹. **c**, The fibrillar nature of the large dense plaques in **b**, shown by negative-stain electron microscopy of purified plaques. (Images **a** and **b** courtesy of B. Chesebro, Rocky Mountain Laboratories; image **c** courtesy of V. Sim, Rocky Mountain Laboratories).

accompanying clinical disease suggests that the lesions caused by anchorless PrPres amyloid are much less problematic than those caused by wild-type PrPres. Surprisingly, in mice that express both the anchorless PrP^C and the GPI-anchored wild-type PrP^C, the presence of anchorless PrP^C accelerates, rather than delays, the onset of clinical disease. This observation suggests that GPI-anchored PrP is important in manifesting the clinically relevant pathological effects of extracellular PrPres. This is consistent with earlier studies that described reduced PrPres-induced neuropathology in brain tissue that lacks PrP^C (refs 20, 21).

Direct or indirect neurotoxic mechanisms

The involvement of GPI-anchored PrP in TSEs is consistent with both direct and indirect mechanisms of PrPres-associated TSE pathogenesis. One possibility is that PrPres is directly toxic to neurons^{22,23} if produced within them in a GPI-anchored state (Box 1), but much less so if it accumulates extracellularly or in non-neuronal cells²¹. However, this does not always seem to be the case, because TSE disease can occur in transgenic mice that express PrP^C only in astrocytes^{24,25}. In addition, there is evidence that exogenous PrPres can induce the release of chemokines such as RANTES that recruit microglia²⁶, which, in turn, could secrete neurotoxic factors (Box 1). A second possibility is that the most neurotoxic entity is not PrPres itself, but an intermediate in, or a labile byproduct of, the conversion pathway between PrP^C and PrPres that must be generated in neuronal membranes to exert its effect¹⁸. A third possibility is that clinical disease is due primarily to the corruption or subversion of the normal functions of PrP^C (ref. 27). Because transgenic PrP^{0/0} mice, which either lack PrP altogether or have their PrP^C expression turned off in adulthood, do not exhibit the symptoms of TSE disease^{20,28,29}, it is unlikely that pathogenesis is simply due to a loss of PrP^C function. Instead, there may be a toxic gain-of-function resulting not only from the presence of PrPres, but from its interactions with, or influence on, PrP^C. Studies of transgenic mice expressing various mutant PrP^C molecules have revealed that perturbations in the structure of PrP^C, its expression levels and functions can cause neurological disease, even in the absence of TSE infection (reviewed in ref. 30). It follows, therefore, that the accumulation of PrPres might also cause disease by altering the metabolism or activities of PrP^C. Thus, in the quest for TSE treatments, it may be important to understand the physiological interactions and functions of PrP^C as well as PrPres.

Diverse ligands and potential functions of PrP^C

Recent studies on the functions of PrP^C suggest that its significance extends well beyond TSE diseases. PrP^C expression is pronounced in neurons but also occurs in various other cell types (Table 1), and diverse activities and ligands have been described for PrP^C (Fig. 2; Box 2; for reviews, see refs 31–34). Particularly notable is the abundant evidence that PrP^C has a role in neuronal neurite^{35,36} (axon³⁷ and dendrite) outgrowth and synapse³⁸ formation.

Despite many interesting revelations about various functions of PrP^C in health and disease, the overall picture remains chaotic. For instance,

it is puzzling, given the importance of many of the proposed functions of PrP, that PrP^{0/0} mice have normal lifespans without obvious problems under standard laboratory conditions (reviewed in refs 28, 29). Although one line of PrP^{0/0} mice exhibited clear-cut neurodegenerative disease, the phenotype proved to be due to ectopic expression of the PrP homologue, doppel, in the brain²⁸. Most phenotypic changes reported in PrP^{0/0} mice have been subtle, perhaps because of backup systems that can at least partly compensate for a lack of PrP^C. Alterations have been reported in circadian rhythm³⁹, hippocampal neuronal function⁴⁰, spatial learning⁴¹, brain copper and cuproenzyme levels⁴², oxidative tissue damage⁴³, phagocytosis and inflammatory response⁴⁴, haematopoietic-stem-cell renewal⁴⁵, neural-stem-cell differentiation⁴⁶ and stress responses⁴⁷. However, some of these phenotypes have been disputed^{48,49}. There have also been claims and counter-claims regarding the modulation of cellular apoptosis by PrP^C (reviewed in ref. 31), the neurotoxicity of PrP peptides^{50,51}, the retrograde transport of PrP^C from the endoplasmic reticulum to the cytoplasm as a key neurotoxic mechanism^{52,53}, and the use of PrP^C as a receptor by *Brucella* species of bacteria^{54,55}.

Common themes for PrP^C functions

Faced with a bewildering array of observations about the functions of PrP^C, it may be useful to consider which, if any, of its known activities might be applicable in the various cell types and physiological processes that are reputed to involve PrP^C. Certain known PrP^C ligands, such as neural cell-adhesion molecule (NCAM), seem to be specific to the nervous system and are not likely to be involved in the function of PrP^C in non-neuronal cells such as leukocytes and haematopoietic stem cells.

PrP^C could have other activities in both neural and non-neuronal cells. For instance, its binding to sulphated glycosaminoglycans (GAGs), such as those in HSPG, and extracellular-matrix proteins, such as laminins and laminin receptors, could be relevant under diverse circumstances because these molecules are widely expressed and serve various functions. In the case of neurons, such interactions might modulate cell adhesion, the homing of neurites, synapse formation and survival. In haematopoietic cells, analogous interactions might facilitate targeting of activated or newly differentiated cells to appropriate tissue sites where their services are needed. One example of this would be the alterations seen in zymosan-induced peritoneal leukocyte recruitment in PrP^{0/0} mice compared with wild-type mice⁴⁴.

The capacity of PrP^C to bind and internalize Cu(II) and Zn(II) (reviewed in refs 33, 34) could also be of functional importance to many types of cell. Although the effect of PrP^C expression on Cu(II) concentrations in the brain is debatable^{42,49}, PrP^C might still serve as a sensor, scavenger or transporter of these metals. Under typical physiological conditions, the affinity of PrP^C for these metal ions is much lower than that of many other Cu(II)- and Zn(II)-binding proteins. However, this affinity can be enhanced by the presence of GAGs⁵⁶.

The restructuring of cells and tissue components is another common feature of the physiological processes in which PrP^C is involved. During

development, learning, cellular differentiation, responses to injury, infections and stresses, some structures are being built (Fig. 2) while others are being cleared away. PrP^C's associations with raft membranes, adhesion molecules and signalling pathways are consistent with its involvement in the assembly of new cellular structures such as neurites and synapses in neurons, and projections in leukocytes (pseudopodia) and follicular dendritic cells (dendritic processes). At the same time, there is evidence that PrP^C negatively modulates the phagocytic activities of macrophages and other phagocytic cells in a manner that might control inflammatory responses to damaged or apoptotic cells⁴⁴. Notably, PrP^C is expressed at relatively high levels in the brain, where overly aggressive inflammatory responses might be particularly damaging and irreversible.

One curious characteristic of PrP^C that is not obviously connected to the assembly of cellular structures but might be related to clearance or scavenging mechanisms is its propensity to bind metals, nucleic acids, porphyrins and their analogues (Box 2). The presence of free forms of these PrP^C ligands in the extracellular spaces would ordinarily be unexpected, unless they were released from pathogens, dying or damaged cells, or perhaps debris from decommissioned cellular structures such as under-used axons, dendritic spines and synapses. The detection and control of potentially harmful factors is an important part of the responses of the innate immune system and glia to infections, tissue damage, stresses, developmental changes (such as apoptosis) and restructuring. It is tempting to speculate that PrP^C might modulate such responses in various cell types by scavenging nucleic acids, metals, porphyrins and similar factors for delivery to appropriate receptors or subcellular sites for degradation and/or signalling. In this context, the binding of PrP^C by cell-surface GAGs might keep PrP^C molecules occupied until preferred ligands appear, because GAGs can compete for the same sites on PrP^C (ref. 57).

Complex formation by PrP^C

Most of the known PrP^C ligands — for example, NCAMs, laminins, laminin receptors and HSPGs — are involved in cellular adhesion, and these molecules, and their binding partners, are notorious for being

polymorphic. Thus, one likely possibility is that PrP^C functions as a part of complexes comprising multiple proteins³² and membrane subdomains that are modulated by cell type, developmental stage, differentiation state, and both extracellular and intracellular cues (Fig. 2). Precise developmental and physiological control of expression of the different polymorphs allows their interactions with PrP^C to be highly context-dependent and to have diverse functional consequences.

PrP^C ligands affecting prion formation

The list of PrP^C-binding molecules is long and growing, and it is important to determine whether these molecules also interact with PrPres, identify their respective interaction sites on PrP^C and PrPres, and characterize the effects of these interactions on PrPres formation. Although the mechanism of PrP^C conversion to PrPres remains to be fully explained, it is clear that specific direct binding between these two molecules precedes the conformational alterations that are associated with the acquisition of protease resistance¹. Unlike most other amyloidogenic proteins and peptides, full-length PrP^C has long been known to be refractory to continuous conversion into PrPres-like structures in most cell-free reactions. This contributed to the hypothesis that cofactors might be required for robust PrP^C conversion to occur. Accessory molecules as varied as chaperones, metal ions, GAGs and lipids can all influence the aggregation of neurotoxic polypeptides. Landmark achievements have recently been made in this area with the identification of cofactors in the form of denaturants and polyanions, such as GAGs and nucleic acids (Box 2), which support indefinite propagation of abnormal PrP aggregates *in vitro*. Importantly, many of these — and related molecules — bind PrP^C, and have been shown to either promote or inhibit PrPres formation both *in vivo* and *ex vivo* (Box 2). In this section, we discuss recent advances in characterizing the role of select PrP^C ligands in the generation of PrPres.

Laminin receptor precursor

The discovery of a positive correlation between the 37-kDa/67-kDa laminin receptor and its precursor (LRP/LR) and PrPres levels suggested that

Table 1 | The cellular distribution and activities of PrP^C in cell types in which known or putative functions have been described

Cell type	Process	Function	Mechanisms, ligands and pathways
Neuron	Neuritogenesis	Adhesion, signalling	Recruits NCAM into rafts to allow it to activate Fyn kinase ³⁵ , which mediates intracellular signalling pathways
			STI-1 binding induces activation of mitogen-activated protein kinase ³⁶
			Binds LRP/LR and HSPG by means of separate sites ⁶⁰
			Binds laminin ⁷⁴
	Synaptogenesis, polarization	Signalling	PrP ^C acts as a growth factor, activating multiple pathways ³⁸
	Survival, trophic effects	Anti-apoptotic	Interacts with BAX ³¹ , STI1 (ref. 36) and NCAM ⁷⁵
		Pro-apoptotic	Binds to anti-apoptotic Bcl-2 (for a review, see ref. 31)
			Crosslinks with anti-PrP antibody ⁷⁶
	Copper binding		Increases levels of p53 (reviewed in ref. 31)
		Copper endocytosis	Induces PrP ^C to aggregate, exit from rafts and undergo clathrin-dependent endocytosis ⁷⁷
Neural stem cells	Redox homeostasis	Copper homeostasis	Maintains appropriate copper levels at the presynaptic membrane and during conditions of oxidative stress (reviewed in ref. 34)
		SOD activity*	Copper-bound PrP ^C has SOD activity (reviewed in ref. 34)
	Neurogenesis	Signalling	Induces NADPH-oxidase dependent ROS through Fyn activation ⁷⁸
	Differentiation	Unknown	Increases cell proliferation in neurogenic regions ⁴⁶
Haematopoietic stem cells	Long-term renewal	Unknown	PrP ^C levels positively influence differentiation ⁴⁶
		Anti-apoptotic? Homing?	Possible mechanisms: transduce cell survival signals; cell-adhesion activity targets cells to appropriate environment; or function as co-receptors for hormones affecting HSC activity ⁴⁵
T cells	Activation	Signalling?	PrP ^C upregulation upon mitogen-induced activation ⁷⁹
	Development	Antioxidant	Copper binding in thymus ⁸⁰
Leukocytes	Differentiation	Unknown	PrP ^C expression by lymphocyte/monocyte lineage ⁸¹
	Phagocytosis	Unknown	PrP ^C modulates phagocytosis ⁴⁴
	Inflammatory response	Homing	PrP ^C alters leukocyte recruitment to site of inflammation ⁴⁴

*Some argue that PrP does not exhibit SOD activity.

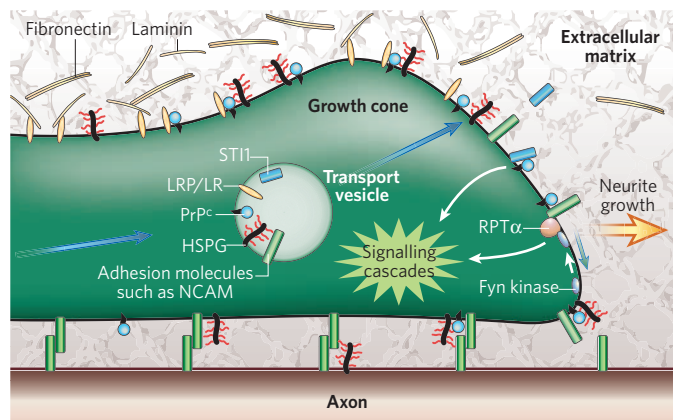


Figure 2 | Model of potential PrP^C interactions associated with axonal growth. PrP^C seems to be important for neurite (nascent axon³⁷ and dendrite) growth and synapse formation³⁸ in neurons. Neurite outgrowth is modulated by PrP^C interactions with NCAM and STI-1, which can lead to activation of intracellular signalling pathways^{35,36}. In the case of NCAM, this signalling pathway is mediated by activation of Fyn kinase, presumably through receptor-type protein phosphatase-α (RTPα). PrP^C binds HSPG^{9–11,71}, laminin⁷⁴ and the laminin receptor (LR) and its precursor (LRP)⁶⁰, which, along with NCAM, are known to mediate contacts between neurons, other cells and the extracellular matrix. Various adhesion and extracellular-matrix molecules help to guide growing neurites to their appropriate destinations. These molecules are presumably delivered to the growing tips (growth cones) of neurites through transport vesicles, perhaps as preformed complexes³².

LRP/LR may participate in PrPres formation⁵⁸. Subsequent experiments in scrapie-infected cell lines have established that LRP/LR is required for PrPres propagation. More recent studies have reported evidence for direct interactions between LRP/LR and PrPres in mediating binding of exogenous PrPres to enterocytes⁵⁹ and baby hamster kidney (BHK) cells⁵⁸, implying that LRP/LR has a role in the initial infection process. A recombinant secreted form of LRP/LR that lacked the transmembrane domain was shown to impair PrPres formation when expressed in scrapie-infected cells, providing a potential therapeutic strategy⁵⁸. Interestingly, direct LRP/LR binding to PrP^C maps to residues 144–179 on PrP^C (ref. 60), overlapping a domain on PrP^C that has been proposed to be involved in binding to PrPres¹. This suggests that dissociation of LRP/LR from PrP^C, perhaps after endocytosis, might be required to permit PrP^C binding to PrPres. It seems likely that LRP/LR serves as a transporter or mediator during conversion, bringing together the two primary substrates for the reaction: PrP^C and PrPres. It will be important to fully characterize the interactions between PrP^C, PrPres, LRP/LR and laminin to determine whether perturbations of binding between these ligands contribute to TSE pathology. Development of LRP/LR knock-down transgenic mice will help to evaluate the role of this protein in prion pathogenesis⁶¹.

Copper

Copper induces the conversion of PrP^C to protease-resistant, detergent-insoluble aggregates^{62,63}, which raises the question of the influence of copper binding on PrPres formation. The effects of copper on PrPres are complex. Copper treatment inhibits PrPres formation in cell-culture models of prion infection and can delay the onset of disease in scrapie-infected hamsters^{64,65}. Paradoxically, copper-chelation therapy also delayed the onset of scrapie in mice⁶⁶. Copper loading can also enhance the protease resistance and recovery of infectivity from partly denatured PrPres preparations^{67,68}. So how can these disparate observations be reconciled? The answer may have been provided by recent studies of cell-free systems of PrP^C aggregation. Copper inhibited PrPres amplification in reactions using purified, brain-derived PrP^C as a substrate and PrPres as a seed⁶⁹. Similarly, spontaneous polymerization of recombinant PrP (rPrP) into amyloid fibrils was inhibited by copper treatment, which

apparently stabilized non-amyloid, α-helical, proteinase-K-resistant rPrP aggregates⁷⁰. Together, these observations might account at least in part for the inhibitory effects of copper on PrPres formation and prion infection. However, when added to pre-formed rPrP fibrils, copper induced coiling and enhanced proteinase K resistance and aggregation⁷⁰. Although the conformation of rPrP in amyloid fibrils seems to differ from that of PrPres generated during prion infection, these observations may help to explain how copper assists regeneration of proteinase-K-resistant PrP in partly denatured PrPres preparations^{67,68}. It remains to be determined whether copper-induced effects on PrP^C trafficking (perhaps to a compartment incompatible with conversion, as depicted in Box 2) or conformation are solely responsible for its inhibitory effect on cellular PrPres formation, although these processes are likely to be interdependent.

Glycosaminoglycans

Sulphated GAGs, particularly those involving heparan sulphate, have long been implicated in interactions between PrPres and PrP^C and TSE disease⁹ (Box 2). In addition to participating in PrPres biosynthesis⁷¹, heparan sulphate was recently identified as a cell-surface receptor for prions^{10,11}. Consistent with observations elsewhere⁷², binding efficiency was independent of host-cell PrP^C expression level, which suggests that PrP^C is not involved in the initial uptake of PrPres as proposed earlier⁷³. As LRP/LR also binds heparan sulphate⁶⁰, and because copper can participate in heparan sulphate binding to PrP^C (ref. 56), it is possible that PrPres interaction with host cells might involve the formation of complexes with all three of the PrP^C ligands highlighted in this section.

Future directions in prion research

Great strides are being made in prion research. As we have discussed, modifications of cell-free PrPres-amplification reactions now permit the perpetual formation of new PrPres, and the first robust cell-free generation of TSE infectivity⁷. Nevertheless, uncertainties still remain about the precise nature of prions. We are eagerly awaiting the results of infectivity studies on cell-free amplified PrPres derived from reactions using purified components⁸, which may unequivocally define the minimal ingredients necessary to generate infectious prions. However, these experiments will not, in themselves, provide critical information about how PrP^C and other potential components become misfolded and assembled into prions. The application of powerful separation and biophysical analysis techniques, such as flow field-flow fractionation, light scattering and nuclear magnetic resonance (NMR), to highly purified PrPres preparations should ultimately provide a complete molecular and biophysical picture of what constitutes the minimal infectious particle^{8,17}. Such studies could also help to elucidate the mechanism(s) of neurotoxicity at work in prion diseases. The most toxic PrPres oligomers — and their mode of action — need to be identified, be it through direct or indirect (or both) effects on the functions of PrP^C. An obvious requirement is an improved understanding of the physiologically relevant function(s) of PrP^C. We hope that future research will answer these questions and promote the development of more effective TSE therapeutics (Box 2).

Although the issues listed above are of great importance, none of these address one of the most intriguing outstanding questions in the field of protein misfolding diseases: why are TSEs transmissible? There is a long list of diseases that are associated with polypeptides that can misfold into amyloid fibrils and other pathological oligomers, but the transmissible nature of TSEs under natural circumstances makes this class of diseases unique. In some cases, a lack of transmissibility might be attributable to intracellular localization of the amyloidogenic protein at a site at which cell-to-cell transfer is difficult. However, many amyloidogenic polypeptides are extracellular. One feature that clearly distinguishes PrP^C from the other amyloidogenic polypeptides is its GPI anchor. This anchor could have numerous effects on the biochemical, biophysical and neurotoxic properties of PrP molecules¹⁹. For instance, GPI anchoring could help to maintain PrPres as a small, membrane-bound oligomeric species — with efficient spreading,

seeding and neurotoxic activities — as opposed to its forming large extracellular amyloid plaques (Fig. 1). Furthermore, there are cellular mechanisms that mediate the exchange of GPI-anchored proteins between cells, including secreted membrane vesicles (such as exosomes and microparticles) and transport via small intercellular projections known as tunnelling nanotubes. These exchange processes could provide a mechanism for the efficient spread of GPI-anchored-protein aggregates to new host cells to seed the formation of new protein aggregates³. It is noteworthy that although TSE infectivity is produced in mice expressing anchorless PrP, it seems to be transmissible only to mice with wild-type, GPI-anchored PrP, which can, in turn, make GPI-anchored PrPres¹⁹. Thus, it is possible that GPI anchoring is key to the unique transmissibility of TSE diseases. ■

- Silveira, J. R., Caughey, B. & Baron, G. S. Prion protein and the molecular features of transmissible spongiform encephalopathy agents. *Curr. Top. Microbiol. Immunol.* **284**, 1–50 (2004).
- Hill, A. F., Antoniou, M. & Collinge, J. Protease-resistant prion protein produced *in vitro* lacks detectable infectivity. *J. Gen. Virol.* **80**, 11–14 (1999).
- Legname, G. et al. Synthetic mammalian prions. *Science* **305**, 673–676 (2004).
- Legname, G. et al. Strain-specified characteristics of mouse synthetic prions. *Proc. Natl Acad. Sci. USA* **102**, 2168–2173 (2005).
- Hsiao, K. K. et al. Serial transmission in rodents of neurodegeneration from transgenic mice expressing mutant prion protein. *Proc. Natl Acad. Sci. USA* **91**, 9126–9130 (1994).
- Nazor, K. E. et al. Immunodetection of disease-associated mutant PrP, which accelerates disease in GSS transgenic mice. *EMBO J.* **24**, 2472–2480 (2005).
- Castilla, J., Saa, P., Hetz, C. & Soto, C. *In vitro* generation of infectious scrapie prions. *Cell* **121**, 195–206 (2005).
- Deleault, N. R. et al. Protease-resistant prion protein amplification reconstituted with partially purified substrates and synthetic polyanions. *J. Biol. Chem.* **280**, 26873–26879 (2005).
- Wong, C. et al. Sulfated glycans and elevated temperature stimulate PrP^{Sc} dependent cell-free formation of protease-resistant prion protein. *EMBO J.* **20**, 377–386 (2001).
- Horonchik, L. et al. Heparan sulfate is a cellular receptor for purified infectious prions. *J. Biol. Chem.* **280**, 17062–17067 (2005).
- Hijazi, N., Kariv-Inbal, Z., Gasset, M. & Gabizon, R. PrP^{Sc} incorporation to cells requires endogenous glycosaminoglycan expression. *J. Biol. Chem.* **280**, 17057–17061 (2005).
- Kovalchuk, O. et al. Cellular heparan sulfate participates in the metabolism of prions. *J. Biol. Chem.* **278**, 40041–40049 (2003).
- Shaked, G. M., Meiner, Z., Avraham, I., Taraboulos, A. & Gabizon, R. Reconstitution of prion infectivity from solubilized protease-resistant PrP and nonprotein components of prion rods. *J. Biol. Chem.* **276**, 14324–14328 (2001).
- Weissmann, C. A unified theory of prion propagation. *Nature* **352**, 679–683 (1991).
- Cordeiro, Y. & Silva, J. L. The hypothesis of the catalytic action of nucleic acid on the conversion of prion protein. *Protein Pept. Lett.* **12**, 251–255 (2005).
- Leucht, C. et al. The 37 kDa/67 kDa laminin receptor is required for PrP^{Sc} propagation in scrapie-infected neuronal cells. *EMBO Rep.* **4**, 290–295 (2003).
- Silveira, J. R. et al. The most infectious prion protein particles. *Nature* **437**, 257–261 (2005).
- Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
- Chesebro, B. et al. Anchorless prion protein results in infectious amyloid disease without clinical scrapie. *Science* **308**, 1435–1439 (2005).
- Brandner, S. et al. Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature* **379**, 339–343 (1996).
- Mallucci, G. et al. Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis. *Science* **302**, 871–874 (2003).
- Kristiansen, M. et al. Disease-related prion protein forms aggregates in neuronal cells leading to caspase activation and apoptosis. *J. Biol. Chem.* **280**, 38851–38861 (2005).
- Hetz, C. et al. The disulfide isomerase Grp58 is a protective factor against prion neurotoxicity. *J. Neurosci.* **25**, 2793–2802 (2005).
- Raeber, A. J. et al. Astrocyte-specific expression of hamster prion protein (PrP) renders PrP knockout mice susceptible to hamster scrapie. *EMBO J.* **16**, 6057–6065 (1997).
- Jeffrey, M., Goodsir, C. M., Race, R. E. & Chesebro, B. Scrapie-specific neuronal lesions are independent of neuronal PrP expression. *Ann. Neurol.* **55**, 781–792 (2004).
- Marella, M. et al. Pathological prion protein exposure switches on neuronal mitogen-activated protein kinase pathway resulting in microglia recruitment. *J. Biol. Chem.* **280**, 1529–1534 (2005).
- Harris, D. A. & True, H. L. New insights into prion structure and toxicity. *Neuron* **50**, 353–357 (2006).
- Hetz, C., Maundrell, K. & Soto, C. Is loss of function of the prion protein the cause of prion disorders? *Trends Mol. Med.* **9**, 237–243 (2003).
- Mallucci, G. & Collinge, J. Rational targeting for prion therapeutics. *Nature Rev. Neurosci.* **6**, 23–34 (2005).
- Chiesa, R. & Harris, D. A. Prion diseases: what is the neurotoxic molecule? *Neurobiol. Dis.* **8**, 743–763 (2001).
- Roucou, X. & LeBlanc, A. C. Cellular prion protein neuroprotective function: implications in prion diseases. *J. Mol. Med.* **83**, 3–11 (2005).
- Martins, V. R. et al. Cellular prion protein: on the road for functions. *FEBS Lett.* **512**, 25–28 (2002).
- Watt, N. T. & Hooper, N. M. The prion protein and neuronal zinc homeostasis. *Trends Biochem. Sci.* **28**, 406–410 (2003).
- Vassallo, N. & Herms, J. Cellular prion protein function in copper homeostasis and redox signalling at the synapse. *J. Neurochem.* **86**, 538–544 (2003).
- Santuccione, A., Sytnyk, V., Leshchynska, I. & Schachner, M. Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth. *J. Cell Biol.* **169**, 341–354 (2005).
- Lopes, M. H. et al. Interaction of cellular prion and stress-inducible protein 1 promotes neurogenesis and neuroprotection by distinct signaling pathways. *J. Neurosci.* **25**, 11330–11339 (2005).
- Moya, K. L., Hassig, R., Breen, K. C., Volland, H. & Di, G. L. Axonal transport of the cellular prion protein is increased during axon regeneration. *J. Neurochem.* **92**, 1044–1053 (2005).
- Kanaani, J., Prusiner, S. B., Diacovo, J., Baekkeskov, S. & Legname, G. Recombinant prion protein induces rapid polarization and development of synapses in embryonic rat hippocampal neurons *in vitro*. *J. Neurochem.* **95**, 1373–1386 (2005).
- Tobler, I. et al. Altered circadian activity rhythms and sleep in mice devoid of prion protein. *Nature* **380**, 639–642 (1996).
- Collinge, J. et al. Prion protein is necessary for normal synaptic function. *Nature* **370**, 295–297 (1994).
- Criado, J. R. et al. Mice devoid of prion protein have cognitive deficits that are rescued by reconstitution of PrP in neurons. *Neurobiol. Dis.* **19**, 255–265 (2005).
- Brown, D. R. et al. The cellular prion protein binds copper *in vivo*. *Nature* **390**, 684–687 (1997).
- Shyu, W. C. et al. Overexpression of PrP^C by adenovirus-mediated gene targeting reduces ischemic injury in a stroke rat model. *J. Neurosci.* **25**, 8967–8977 (2005).
- de Almeida, C. J. et al. The cellular prion protein modulates phagocytosis and inflammatory response. *J. Leukoc. Biol.* **77**, 238–246 (2005).
- Zhang, C. C., Steele, A. D., Lindquist, S. & Lodish, H. F. Prion protein is expressed on long-term repopulating hematopoietic stem cells and is important for their self-renewal. *Proc. Natl Acad. Sci. USA* **103**, 2184–2189 (2006).
- Steele, A. D., Emsley, J. G., Ozdinler, P. H., Lindquist, S. & Macklis, J. D. Prion protein (PrP^C) positively regulates neural precursor proliferation during developmental and adult mammalian neurogenesis. *Proc. Natl Acad. Sci. USA* **103**, 3416–3421 (2006).
- Nico, P. B. et al. Altered behavioural response to acute stress in mice lacking cellular prion protein. *Behav. Brain Res.* **162**, 173–181 (2005).
- Lledo, P. M., Tremblay, P., DeArmond, S. J., Prusiner, S. B. & Nicoll, R. A. Mice deficient for prion protein exhibit normal neuronal excitability and synaptic transmission in the hippocampus. *Proc. Natl Acad. Sci. USA* **93**, 2403–2407 (1996).
- Waggoner, D. J. et al. Brain copper content and cuproenzyme activity do not vary with prion protein expression level. *J. Biol. Chem.* **275**, 7455–7458 (2000).
- Brown, D. R., Schmidt, B. & Kretzschmar, H. A. Role of microglia and host prion protein in neurotoxicity of a prion protein fragment. *Nature* **380**, 345–347 (1996).
- Kunz, B., Sandmeier, E. & Christen, P. Neurotoxicity of prion peptide 106–126 not confirmed. *FEBS Lett.* **458**, 65–68 (1999).
- Ma, J., Wollmann, R. & Lindquist, S. Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**, 1781–1785 (2002).
- Drisaldi, B. et al. Mutant PrP is delayed in its exit from the endoplasmic reticulum, but neither wild-type nor mutant PrP undergoes retrotranslocation prior to proteasomal degradation. *J. Biol. Chem.* **278**, 21732–21743 (2003).
- Watarai, M. et al. Cellular prion protein promotes *Brucella* infection into macrophages. *J. Exp. Med.* **198**, 5–17 (2003).
- Fontes, P. et al. Absence of evidence for the participation of the macrophage cellular prion protein in infection with *Brucella suis*. *Infect. Immun.* **73**, 6229–6236 (2005).
- Gonzalez-Iglesias, R. et al. Prion protein interaction with glycosaminoglycan occurs with the formation of oligomeric complexes stabilized by Cu(II) bridges. *J. Mol. Biol.* **319**, 527–540 (2002).
- Kocisko, D. A. et al. Potent antiscrapie activities of degenerate phosphorothioate oligonucleotides. *Antimicrob. Agents Chemother.* **50**, 1034–1044 (2006).
- Vana, K. & Weiss, S. A Trans-dominant negative 37kDa/67kDa laminin receptor mutant impairs PrP^{Sc} propagation in scrapie-infected neuronal cells. *J. Mol. Biol.* **358**, 57–66 (2006).
- Morel, E. et al. Bovine prion is endocytosed by human enterocytes via the 37 kDa/67 kDa laminin receptor. *Am. J. Pathol.* **167**, 1033–1042 (2005).
- Hundt, C. et al. Identification of interaction domains of the prion protein with its 37-kDa/67-kDa laminin receptor. *EMBO J.* **20**, 5876–5886 (2001).
- Leucht, C. et al. Knock-down of the 37-kDa/67-kDa laminin receptor in mouse brain by transgenic expression of specific antisense LRP RNA. *Transgenic Res.* **13**, 81–85 (2004).
- Kuczius, T. et al. Cellular prion protein acquires resistance to proteolytic degradation following copper ion binding. *Biol. Chem.* **385**, 739–747 (2004).
- Quaglio, E., Chiesa, R. & Harris, D. A. Copper converts the cellular prion protein into a protease-resistant species that is distinct from the scrapie isoform. *J. Biol. Chem.* **276**, 11432–11438 (2001).
- Hijazi, N., Shaked, Y., Rosenmann, H., Ben-Hur, T. & Gabizon, R. Copper binding to PrP^C may inhibit prion disease propagation. *Brain Res.* **993**, 192–200 (2003).
- Kiachopoulos, S., Hesse, J., Tatzelt, J. & Winkhofer, K. F. Misfolding of the prion protein at the plasma membrane induces endocytosis, intracellular retention and degradation. *Traffic* **5**, 426–436 (2004).
- Sigurdsson, E. M. et al. Copper chelation delays the onset of prion disease. *J. Biol. Chem.* **278**, 46199–46202 (2003).
- McKenzie, D. et al. Reversibility of scrapie inactivation is enhanced by copper. *J. Biol. Chem.* **273**, 25545–25547 (1998).
- Nishina, K., Jenks, S. & Supattapone, S. Ionic strength and transition metals control PrP^{Sc} protease resistance and conversion-inducing activity. *J. Biol. Chem.* **279**, 40788–40794 (2004).
- Orem, N. R., Geoghegan, J. C., Deleault, N. R., Kascak, R. & Supattapone, S. Copper (II) ions potently inhibit purified PrPres amplification. *J. Neurochem.* **96**, 1409–1415 (2006).
- Bocharova, O. V., Breydo, L., Salnikov, V. V. & Baskakov, I. V. Copper(II) inhibits *in vitro* conversion of prion protein into amyloid fibrils. *Biochemistry* **44**, 6776–6787 (2005).
- Ben-Zaken, O. et al. Cellular heparan sulfate participates in the metabolism of prions. *J. Biol. Chem.* **278**, 40041–40049 (2003).

72. Magalhaes, A. C. *et al.* Uptake and neuritic transport of scrapie prion protein coincident with infection of neuronal cells. *J. Neurosci.* **25**, 5207–5216 (2005).
73. Baron, G. S., Wehrly, K., Dorward, D. W., Chesebro, B. & Caughey, B. Conversion of raft associated prion protein to the protease-resistant state requires insertion of PrP^{Sc} into contiguous membranes. *EMBO J.* **21**, 1031–1040 (2002).
74. Graner, E. *et al.* Cellular prion protein binds laminin and mediates neuritogenesis. *Brain Res. Mol. Brain Res.* **76**, 85–92 (2000).
75. Chen, S., Mange, A., Dong, L., Lehmann, S. & Schachner, M. Prion protein as trans-interacting partner for neurons is involved in neurite outgrowth and neuronal survival. *Mol. Cell. Neurosci.* **22**, 227–233 (2003).
76. Solforosi, L. *et al.* Cross-linking cellular prion protein triggers neuronal apoptosis *in vivo*. *Science* **303**, 1514–1516 (2004).
77. Taylor, D. R., Watt, N. T., Perera, W. S. & Hooper, N. M. Assigning functions to distinct regions of the N-terminus of the prion protein that are involved in its copper-stimulated, clathrin-dependent endocytosis. *J. Cell Sci.* **118**, 5141–5153 (2005).
78. Schneider, B. *et al.* NADPH oxidase and extracellular regulated kinases 1/2 are targets of prion protein signaling in neuronal and nonneuronal cells. *Proc. Natl Acad. Sci. USA* **100**, 13326–13331 (2003).
79. Cashman, N. R. *et al.* Cellular isoform of the scrapie agent protein participates in lymphocyte activation. *Cell* **61**, 185–192 (1990).
80. Jouvin-Marche, E. *et al.* Overexpression of cellular prion protein induces an antioxidant environment altering T cell development in the thymus. *J. Immunol.* **176**, 3490–3497 (2006).
81. Dodelet, V. C. & Cashman, N. R. Prion protein expression in human leukocyte differentiation. *Blood* **91**, 1556–1561 (1998).
82. Caughey, B. & Raymond, G. J. The scrapie-associated form of PrP is made from a cell surface precursor that is both protease- and phospholipase-sensitive. *J. Biol. Chem.* **266**, 18217–18223 (1991).
83. Borchelt, D. R., Taraboulos, A. & Prusiner, S. B. Evidence for synthesis of scrapie prion protein in the endocytic pathway. *J. Biol. Chem.* **267**, 16188–16199 (1992).
84. Campana, V., Sarnataro, D. & Zurzolo, C. The highways and byways of prion protein trafficking. *Trends Cell Biol.* **15**, 102–111 (2005).
85. Caughey, B., Raymond, G. J., Ernst, D. & Race, R. E. N-terminal truncation of the scrapie-associated form of PrP by lysosomal protease(s): implications regarding the site of conversion of PrP to the protease-resistant state. *J. Virol.* **65**, 6597–6603 (1991).
86. Uehara, T. *et al.* S-nitrosylated protein–disulphide isomerase links protein misfolding to neurodegeneration. *Nature* **441**, 513–517 (2006).
87. Fevrier, B. *et al.* Cells release prions in association with exosomes. *Proc. Natl Acad. Sci. USA* **101**, 9683–9688 (2004).
88. Kocisko, D. A. *et al.* A porphyrin increases survival time of mice after intracerebral prion infection. *Antimicrob. Agents Chemother.* **50**, 759–761 (2006).
89. Raymond, G. J. *et al.* Inhibition of protease-resistant prion protein formation in a transformed deer cell line infected with chronic wasting disease. *J. Virol.* **80**, 596–604 (2006).
90. Cashman, N. R. & Caughey, B. Prion diseases — close to effective therapy? *Nature Rev. Drug Discov.* **3**, 874–884 (2004).
91. Doh-ura, K. *et al.* Treatment of transmissible spongiform encephalopathy by intraventricular drug infusion in animal models. *J. Virol.* **78**, 4999–5006 (2004).
92. Caughey, B. *et al.* Prions and spongiform encephalopathy (TSE) chemotherapeutics: a common mechanism for anti-TSE compounds? *Acc. Chem. Res.* **39**, 646–653 (2006).
93. Caughey, B., Brown, K., Raymond, G. J., Katzenstien, G. E. & Thresher, W. Binding of the protease-sensitive form of PrP (prion protein) to sulfated glycosaminoglycan and Congo red. *J. Virol.* **68**, 2135–2141 (1994).
94. Warner, R. G., Hundt, C., Weiss, S. & Turnbull, J. E. Identification of the heparan sulfate binding sites in the cellular prion protein. *J. Biol. Chem.* **277**, 18421–18430 (2002).
95. Shyng, S. L., Lehmann, S., Moulder, K. L. & Harris, D. A. Sulfated glycans stimulate endocytosis of the cellular isoform of the prion protein, PrP^C, in cultured cells. *J. Biol. Chem.* **270**, 30221–30229 (1995).
96. Yin, S. *et al.* Prion proteins with insertion mutations have altered N-terminal conformation and increased ligand binding activity and are more susceptible to oxidative attack. *J. Biol. Chem.* **281**, 10698–10705 (2006).

Acknowledgements We gratefully acknowledge the helpful comments of W. Caughey, U. Dittmer, B. Chesebro, S. Priola, V. Sim, D. Kocisko, K. Sun Lee and J. Silveira. This work was supported by the Intramural Program of NIAID, NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to the authors (bcaughey@nih.gov; gbaron@nih.gov).

Evolution of the continental crust

C. J. Hawkesworth¹ & A. I. S. Kemp^{1,2}

The continental crust covers nearly a third of the Earth's surface. It is buoyant—being less dense than the crust under the surrounding oceans—and is compositionally evolved, dominating the Earth's budget for those elements that preferentially partition into silicate liquid during mantle melting. Models for the differentiation of the continental crust can provide insights into how and when it was formed, and can be used to show that the composition of the basaltic protolith to the continental crust is similar to that of the average lower crust. From the late Archaean to late Proterozoic eras (some 3–1 billion years ago), much of the continental crust appears to have been generated in pulses of relatively rapid growth. Reconciling the sedimentary and igneous records for crustal evolution indicates that it may take up to one billion years for new crust to dominate the sedimentary record. Combining models for the differentiation of the crust and the residence time of elements in the upper crust indicates that the average rate of crust formation is some 2–3 times higher than most previous estimates.

The Earth's continental crust differs from the crust of other planets in our Solar System. Its formation modified the composition of the mantle and the atmosphere, it supports life, and it remains a sink for CO₂ through weathering and erosion^{1,2}. The continental crust has therefore had a key role in the evolution of this planet, and yet when and how it formed remain the topic of considerable debate. Significant obstacles to understanding this evolution include the shortage of rocks from the first billion years of Earth history and that fact that rocks amenable to study on the surface are very different from those that were involved in the generation of new crust. Thus new techniques have been required to model the differentiation and evolution of the continental crust, and to see back through time to the processes involved in its generation.

The broad compositional features of the continental crust are now well established and have been derived using a number of approaches³. Less well understood are when and how such continental crust was generated, and how the rates and processes involved in the generation of the crust have changed with time. Overall, it seems likely that the rates at which the crust has been generated have decreased with time, as less heat is generated inside the Earth from the decay of radioactive elements. There is a long-standing debate, however, over whether the volume of continental crust has remained essentially unchanged since the earliest Archaean^{4,5}, when it might have been more readily destroyed, or whether the volume of crust has grown through time^{6–12}. Evaluating the composition of new continental crust can provide important clues as to how and when it may have been generated. To do this in turn requires an understanding of the differentiation processes involved in the generation of the compositionally evolved igneous (granites) and sedimentary rocks that dominate the upper crust, and hence the geological record. Although such issues have been historically problematic, they have been reinvigorated by the development of powerful new techniques for the *in situ* analysis of stable and radiogenic isotopes, leading to more robust models for the evolution of the continental crust. Here we highlight these developments and others that have taken place over the past decade¹³. Our emphasis is on the dynamics of crust generation and its evolution, and the extent to which understanding the rates of crust generation influences models for its formation.

The composition of the continental crust

Although the continental crust represents only 0.57% of the mass of the Earth's mantle, it contains, for example, over 40% of elements such as potassium. The bulk continental crust has 60.6% SiO₂ and 4.7% MgO (Table 1), a composition not likely to have been in equilibrium with the upper mantle. Most models for the generation of the continental crust therefore involve at least two stages of differentiation—the extraction of basaltic magma from the mantle and the remelting or fractional crystallization of that basalt, possibly augmented by sedimentary processes^{3,13–19}.

The geological record is dominated by rocks of the upper continental crust, and so models for the formation of new crust require an understanding of how the composition of the upper crust is related to that of new continental crust. The upper crust has a relatively low Eu abundance because Eu partitions strongly into plagioclase feldspar (see, for example, ref. 17), a common residual mineral during crustal melting. The median of minor- and trace-element analyses of granitic rocks with the relative Eu content of the average upper crust is very similar to those in the upper crust¹⁹. The implication is that differentiation of the continental crust is dominated by igneous processes, and that the composition of the upper crust has not been modified significantly by erosion and sedimentation. This knowledge in turn provides the basis for estimating the time-integrated composition of new material added to the continental crust.

There is some debate over the amounts of igneous differentiation that take place below and above the formal base of the crust at the Mohorovičić discontinuity^{16,20}. However, generation of continental crust ultimately involves a basaltic precursor, and for the sake of this discussion we shall start with such basaltic compositions and assume that new material added to the continental crust is typically basaltic in composition¹³. In the modern Earth, basaltic magmas are most voluminously emplaced along mid-ocean ridge systems to form the oceanic crust, but these rocks are unlikely to make a significant material contribution to average continental crust. More relevant are the basalts that form in response to subduction of oceanic crust along destructive plate margins at island or continental arcs, and basalts that are generated in intra-plate settings owing to extensional tectonics, and/or to the emplacement of mantle plumes. The wider implication is the extent to which continental crust is generated in

¹Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK. ²School of Earth and Environmental Sciences, James Cook University, Townsville, Queensland 4811, Australia.

Table 1 | Average and modelled compositions of the continental crust

	Average continental crust ³	Average upper continental crust ³	Upper crust magma composition*	Model new crust ¹⁹ (OIB-arc mix)	Residue ¹⁹
Compositions (wt%)					
SiO ₂	60.60	66.60	68.02		
TiO ₂	0.72	0.64	0.48	0.89	0.95
Al ₂ O ₃	15.90	15.40	15.41		
FeO _t	6.70	5.04	3.30		
MnO	0.10	0.10	0.07		
MgO	4.66	2.48	1.63		
CaO	6.40	3.59	3.62		
Na ₂ O	3.07	3.27	3.60		
K ₂ O	1.81	2.80	2.76	0.61	0.27
P ₂ O ₅	0.13	0.15	0.14	0.15	0.16
Compositions (p.p.m.)					
Rb	49	82	80	10.9	0.1
Ba	456	628	568	212	156
Sr	320	320	301	347	354
Nb	8	12	10.3	4.8	3.9
Y	19	21	21	18.0	17.6
Zr	132	193	157	58.4	42.8
Hf	3.7	5.3	ND	1.6	ND
La	20	31	34	7.8	3.7
Ce	43	63	64	17.9	10.6
Nd	20	27	26	11.0	8.6
Sm	3.9	4.7	5.0	2.9	2.5
Eu	1.1	1.0	1.0	1.0	1.0
Gd	3.7	4	4.1	2.8	2.6
Tb	0.6	0.7	ND	0.5	ND
Dy	3.6	3.9	ND	3.0	ND
Er	2.1	2.3	ND	1.8	ND
Yb	1.9	2	1.9	1.7	1.7
Lu	0.3	0.31	ND	0.3	ND
Pb	11	17	12	2.7	1.3
Th	5.6	10.5	9.3	1.5	0.2
U	1.3	2.7	2.1	0.4	0.1
Ratios					
Eu/Eu*	0.93	0.72	0.70	1.06	1.18
ASI	0.85	1.03	0.99	ND	ND
Mg#	55.3	46.7	0.47	ND	ND

The composition of the average continental crust and the average upper continental crust³, and the upper crust magma composition calculated from global granite data arrays at Eu/Eu* = 0.7 (ref. 19). Also presented is the proposed composition of model new continental crust (itself calculated as a mixture of 92% average arc and 8% OIB, Fig. 1); this has higher U and Th contents than the average lower crust¹⁹, although these are broadly consistent with continental heat flow data⁷⁶. The residue composition is that formed assuming that the *D* values for Rb and Th are ~0 during the differentiation of the continental crust, whereupon the upper crust magma composition reflects 14% melting (or 86% crystallization) of the newly added crust¹⁹. (ASI, alumina saturation index; Mg#, magnesium number; Eu/Eu*, europium anomaly, calculated as in ref. 17; ND, not determined.)

response to either deep-seated thermal anomalies within the Earth (as reflected in mantle plumes) or shallow-level plate-tectonic processes. This can be addressed by chemical modelling, as subduction-related basalts and intra-plate basalts typically have contrasting trace-element compositions. A commonly used diagram is Nb/La plotted against Sr/Nd; the former reflects the distinctive negative mantle normalized Nb anomaly of the continental crust (see, for example, refs 13, 21), whereas the latter ratio is elevated in island arc magmas and it fractionates within the crust because Sr is partitioned into plagioclase feldspar (Fig. 1). Plank¹⁸ highlighted that element ratios that are changed by processes within the continental crust cannot be simply used to estimate the proportions of different materials being added to the crust. Nb/La is slightly lower in the average upper crust than in the average lower crust composition (Fig. 1a), but it appears not to fractionate significantly with increasing silica in crustal rock suites, and so it may be acceptable to use it in this way.

It is difficult to constrain the composition of the basaltic end-member magmas early in the Earth's history. Nevertheless, at the present day intra-plate magmas are reasonably represented by average ocean island basalts (OIB)²² whereas the subduction end-member can be taken as an average of modern primitive island arc basalts (IAB) that contain relatively little contribution from subducted sediment¹⁹. Arc rocks that include significant contributions from subducted sediment^{23,24} have been excluded, because we are concerned with the composition of new material being added to the continental crust, rather than the recycling of pre-existing crustal materials.

Thus the average basaltic protolith to the continental crust (the model new continental crust) should lie at the intersection of arrays between the compositions of magmas generated in subduction and intra-plate settings, and trends for differentiation within the crust (Fig. 1a). It is convenient to calculate that composition in terms of the OIB and IAB end-members (Table 1), but their apparent proportions are much less significant than the proposal that the model new continental crust lies at the intersection of the two arrays. As highlighted by Rudnick¹³, the striking conclusion from Fig. 1 is that the bulk continental crust cannot be modelled as a simple mixture between global average OIB and IAB end-members. This agrees with the major-element evidence that the bulk continental crust represents a fractionated, rather than primary, composition. Instead, the composition of model new continental crust coincides with the estimate of average lower continental crust (Fig. 1a and Table 1).

The composition of the lower continental crust is largely estimated by indirect means^{3,25}, and there is debate over the extent to which lower crustal rocks are residual after melt extraction, and/or have been modified by granulite facies metamorphism. However, Fig. 1 implies that, on average, the lower crust is similar in composition to unmodified basalts, and this conclusion is insensitive to the choice of different within-plate basalt end-members. It has therefore been inferred¹⁹ that the average lower crust is broadly representative of the protolith to the continents.

It has been proposed, and it could be inferred from Fig. 1a, that the model new continental crust may be simply represented as a mixture

of ~8% OIB and ~92% IAB^{13,21}. This would in turn appear to reinforce the importance of convergent margin processes in the generation of continental crust^{14,16}. However, such figures should be regarded with caution. For example, in the Archaean, significant volumes of new crust were associated with the emplacement of the tonalite-trondhjemite-granodiorite (TTG) intrusive suites that are characteristic of this time period^{17,26}. The TTG are thought to have been generated by partial melting of hydrated basalt²⁷, and in such melting processes low Nb/La ratios are associated with low Ti/Zr ratios (see, for example, ref. 17). This differs from recent subduction settings where tholeiitic basalts with low Nb/La are commonly generated without an associated fractionation of Ti. The significance is that in the Archaean, low Nb/La ratios could have resulted from remelting basaltic source rocks even if those rocks were generated in a within-plate setting (Fig. 1a). This highlights the potential pitfalls in assessing how the proportions of new crust generated in different settings may have changed with time²⁸.

Differentiation of the continental crust

Differentiation of the continental crust is dominated by igneous processes¹⁹. The degree of differentiation constrains how the rates at which the upper crust is generated relate to those at which new crust is generated, and the amounts of residual crustal material recycled back into the mantle. The degree of differentiation can be estimated from the compositions of model new crust and the upper continental crust¹⁹. Calculations reveal that the upper crust reflects ~14% partial melting, or the analogous amount of fractional crystallization, of average new basaltic crust (Table 1). Inevitably, the residual composition is very different from the average lower crust in that it has strongly depleted incompatible-element abundances (for example, U, Th, Pb, La, Zr) and relatively high Ti contents; in other words, it is more refractory and genuinely residual. Relative depletions in certain elements during crustal differentiation reflect the principal mineralogical controls in the differentiation of the continental crust. Specifically, Eu and Ti highlight the role of residual plagioclase and a Ti-rich phase, and there is little evidence for residual garnet during crust differentiation, which would lead to lower contents of Y and heavy rare-earth elements. This is consistent with Nd-Hf isotope studies on the lower crust (see, for example, refs 29 and 30), highlighting that most of the continental crust differentiated at depths that were too shallow for significant amounts of residual garnet. Differentiation, and perhaps the generation of significant volumes of new crust, does not appear to have been associated with areas of thickened continental crust.

A consequence of the estimated proportions of new and upper crustal material is the large volume of residual material after differentiation. The upper crust is ~12.5 km thick in crust with an average total thickness of 40 km (ref. 3). If the upper crust is the product of 14% melting, the corresponding residue would be 77 km thick, resulting in a total crustal thickness of ~100 km, including the middle crust. There is no geophysical evidence for a contemporary mafic lower crust of this thickness³, and the appropriate melt-depleted compositions are uncommon in the spectrum of exposed lower crustal lithologies²⁵. Such a thick residual layer would be gravitationally unstable under most geothermal conditions, owing to phase transformations that produce dense minerals like garnet^{31–33}. It is therefore concluded that the volumetrically dominant crustal residue of upper crust formation has largely foundered into the mantle^{16,17,23,31–33}, with the consequence that the residence time of material in the lower crust is much shorter than in the upper crust. This reinforces arguments that delamination of residual lower crustal material is critical for establishing the andesitic composition of the average continental crust (Fig. 1). The residual material returned to the mantle would have had relatively high Sr/Nd and low Rb/Sr, Sm/Nd and Th/La ratios^{13,18,32,34,35}, and it has been invoked as the cause of trace-element-enriched material in the source of some OIB^{35,36}.

If the average residence times of elements in the lower crust are so much less than those in the upper crust, probably at least five times less¹⁹, what are the geological implications? When is material lost from the lower crust? Crustal differentiation is a logical thermal consequence of crust generation, as seen for example in the Andes³⁷ and eastern Australia^{17,19}, particularly where crustal melting is driven by basaltic under/intraplating (see, for example, refs 38–40). Rapid processing of the lower crust, and delamination of the residual material, may therefore largely occur at the sites of active crustal growth^{32,33,41}, and perhaps less frequently in response to lithospheric thickening linked to compressional tectonics⁴².

Reconciling the sedimentary and igneous records

Fine-grained detrital sedimentary rocks such as shales provide the most representative samples of the upper continental crust. They have been used to determine average compositions for the abundances of at least the insoluble elements in the upper crust^{14,43} and to constrain the Nd isotope evolution of the continental crust^{44–46}.

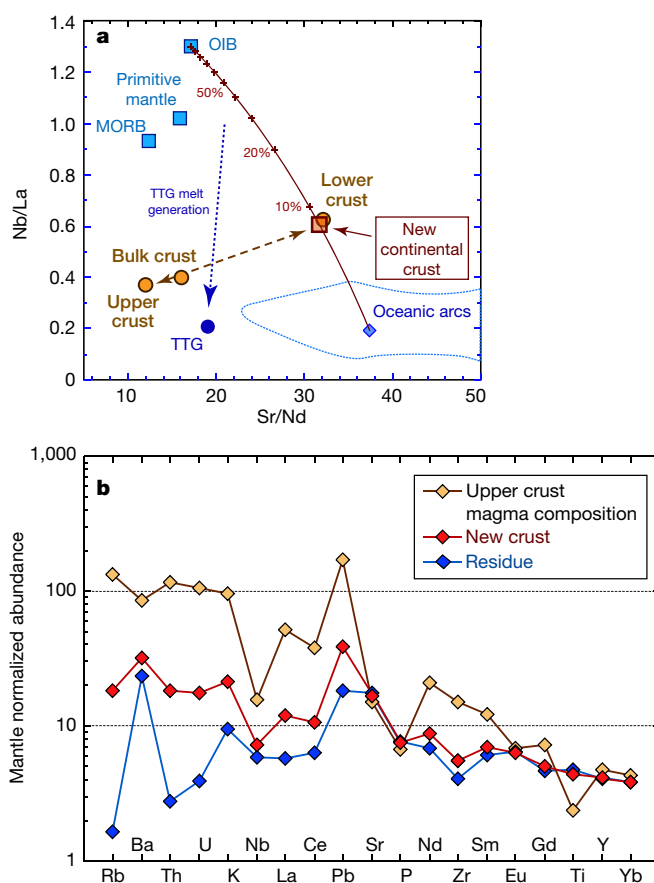


Figure 1 | Trace-element ratios and abundances. **a**, A plot of the trace-element ratios Nb/La versus Sr/Nd, following refs 13 and 19. It illustrates how the estimated composition of model new continental crust is taken to lie at the intersection of the differentiation trend between the bulk continental crust, the upper and the lower crust, and the array between basalts generated in within-plate and subduction-related settings. The composition of model new continental crust is similar to previous estimates for the lower continental crust³. MORB, mid-ocean ridge basalt; OIB, ocean island basalt; and TTG, tonalite-trondhjemite-granodiorite. The steep vector to average TTG illustrates a general model for generating TTG from the remelting of hydrated basalt. **b**, Mantle normalized minor- and trace-element abundances of the estimated composition of model new continental crust and the composition of granites with Eu/Eu* of 0.7 that is very similar to average upper crust³ (Table 1). The complementary residue composition is also shown. Assuming that the distribution coefficients for elements like Rb and Th are close to zero, the upper crust represents 14% partial melt of new crust, or an analogous degree of fractional crystallization¹⁹.

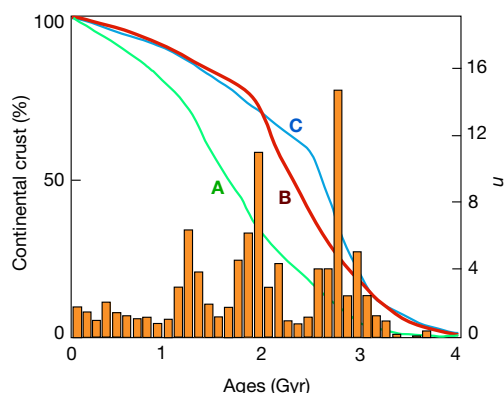


Figure 2 | Generation of continental crust through time as estimated from different approaches. The histogram shows the ages of igneous rocks with mantle-like initial isotope ratios that may therefore be taken to represent the generation of new continental crust (n is the number of samples)⁴⁷. This distribution of ages is contrasted with selected curves for estimates of the increase in the volume of stable continental crust with time. Curve A is based on Nd isotopes in shales⁴⁴, B on Pb isotopes⁷⁷ and C is from ref. 78.

However, it has proven more challenging to reconcile the picture of crustal evolution obtained from the sedimentary reservoir with the complementary information available from the igneous rock record.

Compilations of emplacement ages of isotopically juvenile (mantle-derived) igneous rocks, exposed on different continents, show striking peaks around 2.7, 1.9 and 1.2 Gyr ago^{47,48} (Fig. 2). It is possible that these peaks simply reflect the preservation history of old rocks, but they have been interpreted as periods of accelerated crust formation, each spanning several hundred million years. It is difficult to conceive how global pulses of crustal growth might occur in response to conventional Wilson-style plate tectonics¹⁵, particularly because there are no age peaks in the past billion years when we may surmise that subduction-related magmatism dominated the generation of new crust (Fig. 2). An alternative is that periodicity in igneous activity reflects rapid crust formation during thermal pulses associated with the emplacement of mantle plumes^{15,47–50}. There is geological support for this in the plume affinity of basaltic rocks associated with vast tracts of juvenile crust formed over short time periods, such as in the Birimian (2.1 Gyr ago) terranes of West Africa⁴⁹ and the Arabian-Nubian Shield⁵⁰. More widely, this reaffirms the distinction between crust generated predominantly in response to deep-seated disturbances in the Earth, and that generated by shallow-level processes, as in some form of plate tectonics.

In sharp contrast, models for the volume of the continental crust result in smooth curves through time, with as yet no sign of marked episodicity^{4–12} (Fig. 2). The curves are smooth in part because these models are typically based on radiogenic isotope ratios in sediments and sedimentary rocks. Such samples reflect the history of the crust over long periods of time^{43–46,51}, but they may also result in hybrid crust generation ages that are the result of mixing processes rather than real magmatic events. Igneous rocks preserve the ages of the specific events in which new continental crust may have been generated (as in Fig. 2), although this record is undoubtedly incomplete and may not be representative of the true growth history of the continents. The key therefore is to have an archive in which the evolution of sediments and igneous rocks can be studied separately.

The accessory mineral zircon offers such an archive. Zircons crystallize from medium- to high-silica magmas and are robust under most crustal conditions, being able to survive intense metamorphism and prolonged sedimentary recycling. Detrital zircons in sediments thus encapsulate the geological history of the continental crust^{51–54}, back even as far as 4.4 Gyr ago^{55,56}. Zircons can be dated precisely using U-Pb isotopes⁵⁷, and are amenable to *in situ* analysis of both Hf and O isotopes^{51,58}. Hf isotopes are analogous to Nd isotopes in that they indicate whether a rock was derived from the mantle or from the

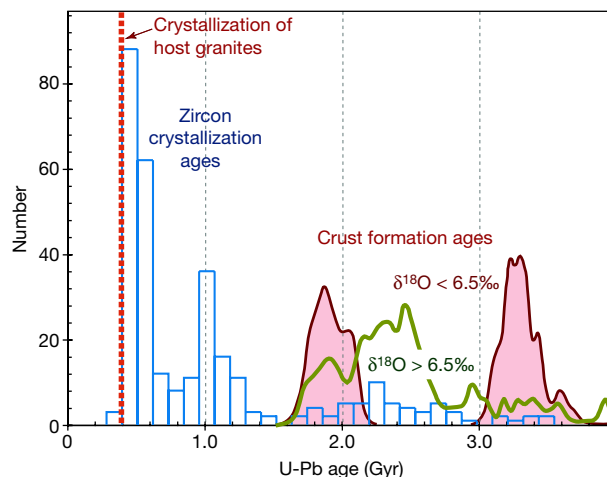


Figure 3 | Histogram of detrital and inherited zircon crystallization ages from southeast Australia overlain with crust formation ages inferred from the Hf-O isotope compositions of the same grains. Data from ref. 66. The thick vertical dashed line shows the crystallization ages of the granitic rocks that contained the inherited zircons. The major magmatic events recorded by the detrital and inherited zircons (blue histogram: ~500 Myr ago and ~1 Gyr ago) were not obviously associated with peaks in the generation of new continental crust, which instead largely occurred in two major episodes at ~1.9 Gyr ago and 3.3 Gyr ago (low- $\delta^{18}\text{O}$ samples). The samples with higher $\delta^{18}\text{O}$ (greenish-brown line) are inferred to have hybrid crust formation ages. $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}} - 1] \times 1,000$; the VSMOW standard is used.

pre-existing crust. However, zircons are less prone to isotope disturbance of the kind that has hampered studies of old bulk rocks^{59,60}, and they preserve $^{176}\text{Hf}/^{177}\text{Hf}$ isotope ratios close to the isotope ratio of the magma at the time of crystallization⁶¹. This initial ratio can then be used to calculate a model Hf crustal residence age, which is the average time at which the source rock(s) of the magma from which the zircon crystallized were derived from the mantle^{52–54,62–66}. These model Hf ages are thus commonly referred to as crust formation ages.

The $\delta^{18}\text{O}$ values of materials recycled at the Earth's surface are higher than those of igneous rocks that have escaped low temperature interactions with the hydrosphere. Thus oxygen isotopes in zircons can be used to distinguish those that crystallized from magmas that contained a contribution from sediments from those derived from infracrustal igneous source rocks^{51,58,65,66} (Fig. 3). The former, those with higher $\delta^{18}\text{O}$ values, may yield hybrid Hf model ages that are difficult to interpret, and hence may not reflect particular crust-forming events. However, the Hf model ages from zircons with low $\delta^{18}\text{O}$ values are thought to record the Hf isotope ratios of magmas derived from igneous precursors, whereupon they are much more likely to reflect particular crust forming events.

This approach has recently been illustrated with detrital and inherited zircons from Palaeozoic rocks of southeast Australia⁶⁶. The main observations of this study are summarized in Fig. 3. There were major magmatic episodes at ~0.5 and 1 Gyr ago, but these did not involve the generation of new crust because these crystallization ages do not coincide with the Hf crust formation ages. The zircons with low $\delta^{18}\text{O}$ are from magmas derived from igneous sources. The Hf model ages of these grains therefore indicate periods of new crustal growth at 1.9 and 3.3 Gyr ago. In contrast, the zircons with high $\delta^{18}\text{O}$ crystallized from magmas with a sedimentary component, and thus may have mixed age source rocks. The Hf model ages of these zircons peak at 2.2–2.6 Gyr ago, and represent a mixture of the rocks generated at 1.9 and 3.3 Gyr ago. The oxygen isotope data allow us to infer that the 2.1–2.6 Gyr age peak represents mixing in the sedimentary environment, rather than a real crust-forming event. More generally, with this approach we can now contrast the evolution of the igneous and

sedimentary reservoirs in the continental crust, and compare the information from zircons with the models for the evolution of the continental crust based on Nd isotope ratios in shales⁴⁴.

Figure 4 summarizes the data from zircons and shales from Australia on a graph of the model Nd or Hf age (that is, the crust formation age for each isotope system), plotted against the sedimentation age or the age of the individual zircons, respectively. The decrease in the model Nd age with the age of sedimentation for the shales indicates that significant volumes of new crust were generated as least until the middle Proterozoic. Models for the evolution of the crust were based on assumed episodes of generation of new crust every 0.5 Gyr, and required this new material to have been rapidly eroded into the sedimentary reservoir⁴⁴. Yet the available data from the ages of rocks (Fig. 2) and zircons (Figs 3 and 4) now indicate that the crust may have been generated in fewer and more widespread events, at least in the period 2.9–1.2 Gyr ago.

On Fig. 4, zircons that crystallized from magmas that were new additions to the crust have the same crystallization and crust formation age, and so plot on the black diagonal line. But very few zircons

actually plot on that line, which may highlight a time difference between generation of new continental crust and the formation of granitic magmas that can crystallize zircon. Once new crust has been formed, zircons that crystallize from magmas generated from that crust plot on a horizontal array, reflecting the ages of subsequent thermal events. It is very striking that evidence for the generation of new crust at 3.3 Gyr ago is retained in zircons that crystallized over the subsequent 2 Gyr (Fig. 4).

As with Fig. 3, the low- $\delta^{18}\text{O}$ zircons in Fig. 4 define the crust-forming events at 1.9 Gyr ago and 3.3 Gyr ago. To reproduce the Nd isotope curve in sediments from only a few crust-forming events implies that the slope of that curve in part reflects the rate at which new crust enters the sedimentary record. Although the data set is still small, the results from the high- $\delta^{18}\text{O}$ zircons also indicate that it may take a long time, perhaps up to ~ 1 Gyr, for new crustal material to dominate the sedimentary record (Fig. 4). Such relative rates of erosion of different aged source terrains are difficult to reconcile with the development and envisaged rapid erosion of mountainous areas, as typically form along destructive plate boundaries. The implication is that much new crust is emplaced within the pre-existing crust by under- or intraplating^{38,67}. This delays its erosion and contribution to the sedimentary record, and it is consistent with significant volumes of crust being generated in within plate settings.

Rates of generation of the continental crust

Radiogenic isotopes constrain the volumes of continental crust that were stable for long enough for the radiogenic isotopes to change in response to radioactive decay, and the rates of increase in the volume of such stable crust. They do not constrain the overall rates of generation of continental crust. The latter have instead been investigated using magma addition rates at volcanic arcs and intra-plate hotspots, and the residence times of elements in the upper crust. The estimated rates at which basaltic magma is added to the continental crust range from $1.65 \text{ km}^3 \text{ yr}^{-1}$ (ref. 68) to $3.7 \text{ km}^3 \text{ yr}^{-1}$ (ref. 69) if destruction of continental crust by forearc erosion and sediment subduction is taken into account⁷⁰. Such data are inevitably restricted to the recent geologic past, and it is widely accepted that the rates of crust generation were higher earlier in the history of the Earth⁷¹.

Another approach links the annual flux of new material into the continental crust to the residence times of elements in the crust

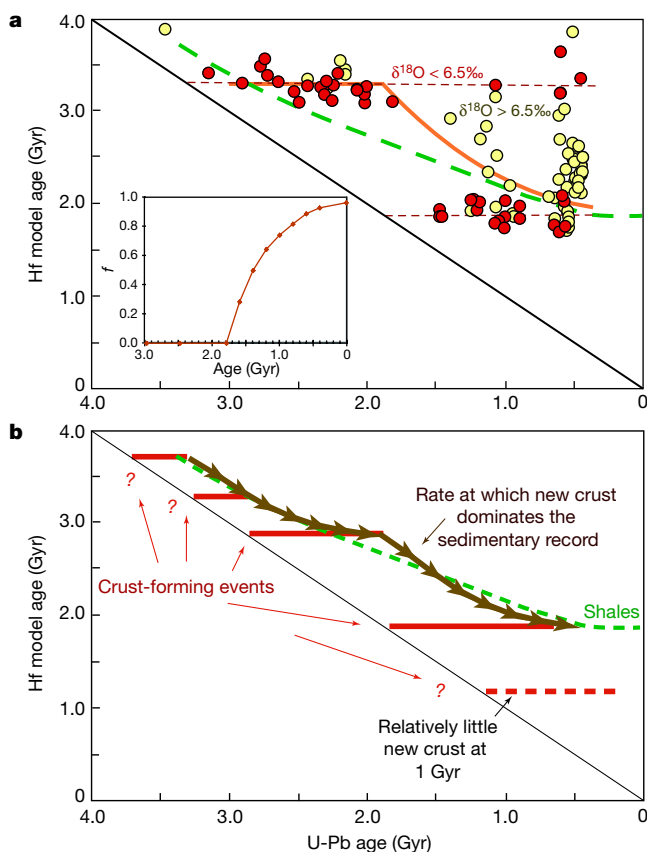


Figure 4 | Hf and U-Pb model ages. **a**, **b**, Hf model ages, or the Hf crust formation ages, plotted against the U-Pb crystallization ages of different zircon suites⁶⁶. The dashed horizontal lines represent particular times of crust formation, and the diagonal black line indicates 1:1 correspondence between zircon U-Pb crystallization ages and the 'model' crust formation ages inferred from Hf isotopes. **a**, The red circles represent low- $\delta^{18}\text{O}$ zircons, and the yellow circles are the high- $\delta^{18}\text{O}$ zircons. The latter are inferred to contain a contribution from sediments, and as such their model ages may be hybrids that reflect mixing rather than specific crust forming events. The green dashed line is from a plot of Nd model age versus sedimentation age for average Australian shale compositions⁴⁴. The inset shows the fraction (f) of 1.9-Gyr crust contributing to the sedimentary reservoir as a function of time, as modelled from the high $\delta^{18}\text{O}$ zircon trend (orange curve). **b**, An alternative interpretation of the evolution of the upper crust, in which new crust can take up to 1 Gyr to dominate the sedimentary record (arrowed brown curve). The solid red bars mark times of crust generation, whereas the dashed red bar represents a magmatic event in which little new crust formed.

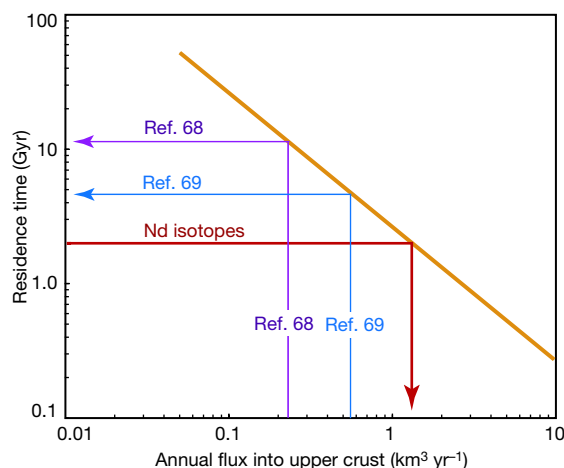


Figure 5 | The relationship between the residence time of elements in the upper crust and the annual rates of crust generation. This was calculated using the present-day volume of continental crust, and assuming that the rate of upper crust generation was 15% of the rate of new crustal addition. For a residence time of 2 Gyr in the upper crust (the model Nd age of the upper crust), the rates of crust generation are >6 times those in the recent geologic past, and 2–3 times greater than the average rates inferred from radiogenic isotopes. Also shown are residence times calculated from crust generation rates obtained from refs 68 and 69.

(Fig. 5)¹⁹. For the rare earth elements, the residence time can be taken to be the average model Nd age of the upper crust, ~2 Gyr (ref. 44). A critical step is to determine how the rates of generation of the upper crust compare with those for the generation of bulk new continental crust. As discussed above, the annual flux into the upper crust is ~14% of the rate at which new crust is generated. Thus the estimated rates of magma addition to the continental crust of 1.65 and 3.7 km³ yr⁻¹ (refs 68, 69) are equivalent to values of 0.23 and 0.56 km³ yr⁻¹ for the rate of formation of new upper crust. These correspond to upper crustal residence times of 11 Gyr and 4.7 Gyr, respectively, which are unrealistic as they exceed the age of the Earth. These calculations highlight how the rates of continental crust generation, and of the fluxes of differentiated magma into the upper crust, must have been greater earlier in Earth history than in the relatively recent geologic past. In contrast, a residence time of 2 Gyr for elements in the upper crust converts to an average rate of generation of 1.3 km³ yr⁻¹ for the upper crust (Fig. 5), and of 9.3 km³ yr⁻¹ for the bulk continental crust. This value is substantially higher than the average rate of crustal growth if the present volume has been generated since 3.5 Gyr ago, which is 2.2 km³ yr⁻¹. But given that crustal material is recycled into the mantle, the true rates of crust generation are inevitably higher than the rates at which volumes of continental crust were stabilized.

Future directions

There is increasing evidence that the rates of generation of the continental crust, and the ways in which it has been generated, have changed with time. The average rates were higher in the past, and the data from the rock record and from zircons also indicate that there were periods of relatively rapid crustal growth, and periods when much less crust was generated (Fig. 2). The implication is that the pulses of rapid crustal growth were linked to thermal instabilities within the Earth, perhaps manifest as superplumes. Such pulses are much less obvious in the past 1 Gyr of Earth history when it is presumed that the generation and reworking of continental crust were more closely associated with plate tectonics and magmatism along destructive plate margins.

Developments in the *in situ* analysis of zircon have made it possible separately to investigate the igneous and sedimentary reservoirs in the continental crust. These also indicate that significant volumes of continental crust were generated in major episodes in the Archaean and the early Proterozoic, and they highlight that long periods of time may be required before the new crust begins to dominate the sedimentary record. This modifies models for erosion of the continental crust. It is argued that the residence times of elements in the upper crust are much greater than in the lower crust. The annual flux of material into the upper crust can in principle be inferred from its volume and the residence times of elements in the upper crust. A maximum value of the latter is provided by the model Nd age of the upper crust of 2 Gyr, and this indicates that the average rates of crust generation (~9.3 km³ yr⁻¹) are in excess of six times those in the recent geologic past, and 2–3 times greater than the rates inferred from radiogenic isotopes. Simple calculations based on this establish that more than half the K, and a quarter of the Li, in the silicate Earth may have been processed through the continental crust over the past 4 Gyr (ref. 19).

Such arguments bring us back to the difficulties in assessing the past volumes of continental crust, particularly in the Archaean. Relict outcrops of rocks containing exceptionally old zircons, such as at Jack Hills in Australia^{55,56}, provide little constraint as there is no indication of how representative these zircons may be. The isotope evidence for depleted upper mantle by 3.8 Gyr ago and possibly 4 Gyr ago^{72,73} indicates that crust had been extracted by that time, and the parent/daughter element ratios are more readily fractionated in the generation of continental rather than basaltic oceanic crust. Widespread depleted mantle therefore offers encouragement to ideas of significant volumes of continental crust, although it may be residual after

mafic crust formation provided that the degree of melting is reasonably small (<10%). Preliminary thermal models indicate that it is difficult to envisage a hotter Earth in which basaltic crust is not remelted to form high-silica crust. New insights may be available from the Nd and Ca isotope ratios in other well dated mineral archives, such as titanite⁷⁴, in the Archaean, and the isotope composition of elements fractionated in the hydrosphere and returned to the mantle, such as Li (ref. 75). The ways forward include improving those models for melt generation in the early Archaean, and determining the distribution of crystallization and crust formation ages as they were at different times in the Archaean from integrated Hf-O isotope studies of detrital and inherited zircons in old rocks.

1. Lowe, D. R. & Tice, M. M. Geologic evidence for Archaean atmospheric and climatic evolution: fluctuating levels of CO₂, CH₄ and O₂, with an overriding tectonic control. *Geology* **32**, 493–496 (2004).
2. Kramers, J. D. Global modelling of continent formation and destruction through geological time and implications for CO₂ drawdown in the Archaean Eon. *Spec. Publ. Geol. Soc. Lond.* **199**, 259–274 (2002).
3. Rudnick, R. L. & Gao, S. in *The Crust* (ed. Rudnick, R. L.) 1–64 (Treatise in Geochemistry, Vol. 3, Elsevier, Amsterdam, 2003).
4. Armstrong, R. L. Radiogenic isotopes: the case for crustal recycling on a near steady state non-continental-growth model. *Phil. Trans. R. Soc. Lond. A* **301**, 443–472 (1981).
5. Fyfe, W. S. The evolution of the Earth's crust: modern plate tectonics to ancient hot spot tectonics? *Chem. Geol.* **23**, 89–114 (1978).
6. Hurley, P. M. & Rand, J. R. Pre-drift continental nuclei. *Science* **164**, 1229–1242 (1969).
7. Allègre, C. J. & Lewin, E. Chemical structure and history of the Earth: evidence from global non-linear inversion of isotopic data in a three-box model. *Earth Planet. Sci. Lett.* **96**, 61–88 (1989).
8. Moorbath, S. Age and isotope evidence for the evolution of the continental crust. *Phil. Trans. R. Soc. Lond. A* **288**, 401–413 (1978).
9. O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. Geochemical modelling of mantle differentiation and crustal growth. *J. Geophys. Res.* **84**, 6091–6101 (1979).
10. Jacobsen, S. & Wasserburg, G. J. The mean age of mantle and crustal reservoirs. *J. Geophys. Res.* **84**, 7411–7427 (1979).
11. Kramers, J. D. & Tolstikhin, I. N. Two major terrestrial Pb isotope paradoxes, forward transport modelling, core formation and the history of the continental crust. *Chem. Geol.* **139**, 75–110 (1997).
12. Albarède, F. & Brouxel, M. The Sm/Nd secular evolution of the continental crust and the depleted mantle. *Earth Planet. Sci. Lett.* **82**, 25–35 (1987).
13. Rudnick, R. L. Making continental crust. *Nature* **378**, 573–578 (1995).
14. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell, Malden, MA, 1985).
15. Albarède, F. The growth of continental crust. *Tectonophysics* **296**, 1–14 (1998).
16. Arculus, R. J. Origins of the continental crust. *J. Proc. R. Soc. New South Wales* **132**, 83–110 (1999).
17. Kemp, A. I. S. & Hawkesworth, C. J. in *The Crust* (ed. Rudnick, R. L.) 349–410 (Treatise in Geochemistry, Vol. 3, Elsevier, Amsterdam, 2003).
18. Plank, T. Constraints from thorium/lanthanum on sediment recycling at subduction zones and the evolution of continents. *J. Petrol.* **46**, 921–944 (2005).
19. Hawkesworth, C. J. & Kemp, A. I. S. The differentiation and rates of generation of the continental crust. *Chem. Geol.* **226**, 134–143 (2006).
20. Turner, S. P. et al. Time-scales of magma formation, ascent and storage beneath subduction-zone volcanoes. *Phil. Trans. R. Soc. Lond. A* **358**, 1443–1464 (2000).
21. Barth, M., McDonough, W. F. & Rudnick, R. L. Tracking the budget of Nb and Ta in the continental crust. *Chem. Geol.* **165**, 197–213 (2000).
22. Sun, S.-s. & McDonough, W. F. Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes. *Spec. Publ. Geol. Soc. Lond.* **42**, 313–345 (1989).
23. Plank, T. & Langmuir, C. H. The geochemical composition of subducting sediment and its consequences for the crust and mantle. *Chem. Geol.* **145**, 325–394 (1998).
24. Hawkesworth, C. J., Turner, S. P., McDermott, F., Peate, D. W. & van Calsteren, P. U-Th isotopes in arc magmas: implications for element transfer from the subducted crust. *Science* **276**, 551–555 (1997).
25. Rudnick, R. L. & Fountain, D. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
26. Martin, H. in *Archaean Crustal Evolution* (ed. Condie, K. C.) 205–260 (Elsevier, Amsterdam, 1994).
27. Rapp, R. P. & Watson, E. B. Dehydration melting of metabasalt at 8–32 kbar: implications for continental growth and crust-mantle recycling. *J. Petrol.* **36**, 891–931 (1995).
28. Condie, K. C. in *Archaean Crustal Evolution* (ed. Condie, K. C.) 85–120 (Elsevier, Amsterdam, 1994).
29. Vervoort, J. D. & Patchett, P. J. Behaviour of hafnium and neodymium isotopes in the crust: constraints from Precambrian crustally derived granites. *Geochim. Cosmochim. Acta* **60**, 3717–3723 (1996).
30. Vervoort, J. D. et al. Hf-Nd isotopic evolution of the lower crust. *Earth Planet. Sci. Lett.* **181**, 115–129 (2000).

31. Kay, R. W. & Kay, S. M. Creation and destruction of lower continental crust. *Geol. Rundsch.* **80**, 259–278 (1991).
32. Kay, R. W. & Kay, S. M. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
33. Jull, M. & Kelemen, P. B. On the conditions for lower crustal convective instability. *J. Geophys. Res.* **106**, 6423–6446 (2001).
34. Ellam, R. M. & Hawkesworth, C. J. Is average continental crust generated at subduction zones? *Geology* **16**, 314–317 (1988).
35. Arndt, N. T. & Goldstein, S. L. An open boundary between lower continental crust and mantle: its role in crust formation and crustal cycling. *Tectonophysics* **161**, 201–212 (1989).
36. Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
37. Petford, N. & Atherton, M. Na-rich partial melts from newly underplated basaltic crust: the Cordillera Blanca Batholith, Peru. *J. Petrol.* **37**, 1491–1521 (1996).
38. Voshage, H. *et al.* Isotopic evidence from the Ivrea Zone for a hybrid lower crust formed by magmatic underplating. *Nature* **347**, 731–736 (1990).
39. Annen, C. & Sparks, R. S. J. Effects of repetitive emplacement of basaltic intrusions on thermal evolution and melt generation in the crust. *Earth Planet. Sci. Lett.* **203**, 937–955 (2002).
40. Bergantz, G. W. & Dawes, R. in *Magmatic Systems* (ed. Ryan, M. P.) 291–317 (Academic, San Diego, CA, 1994).
41. Zandt, G. *et al.* Active foundering of a continental arc root beneath the southern Sierra Nevada in California. *Nature* **431**, 41–46 (2004).
42. Moore, V. M. & Wiltschko, D. V. Syn collisional delamination and tectonic wedge development in convergent orogens. *Tectonics* **23**, doi:10.1029/2002TC001430 (2004).
43. McLennan, S. M. Relationships between the trace element composition of sedimentary rocks and upper continental crust. *Geochim. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000109 (2001).
44. Allègre, C. J. & Rousseau, D. The growth of continents through geological time studied by Nd isotope analysis of shales. *Earth Planet. Sci. Lett.* **67**, 19–34 (1984).
45. O'Nions, R. K., Hamilton, P. J. & Hooker, P. J. A Nd isotope investigation of sediments related to crustal development in the British Isles. *Earth Planet. Sci. Lett.* **63**, 229–240 (1983).
46. Davies, G., Gledhill, A. & Hawkesworth, C. Upper crustal recycling in southern Britain: evidence from Nd and Sr isotopes. *Earth Planet. Sci. Lett.* **75**, 1–12 (1985).
47. McCulloch, M. T. & Bennett, V. C. Progressive growth of the Earth's continental crust and depleted mantle: geochemical constraints. *Geochim. Cosmochim. Acta* **58**, 4717–4738 (1994).
48. Condie, K. C. Episodic continental growth and supercontinents: a mantle avalanche connection? *Earth Planet. Sci. Lett.* **163**, 97–108 (1998).
49. Boher, M., Abouchami, W., Michard, A., Albarede, F. & Arndt, N. T. Crustal growth in West Africa at 2.1 Ga. *J. Geophys. Res.* **97**, 345–369 (1992).
50. Stein, M. & Hofmann, A. W. Mantle plumes and episodic crustal growth. *Nature* **372**, 63–68 (1994).
51. Valley, J. W. *et al.* 4.4 billion years of crustal maturation: oxygen isotope ratios of magmatic zircon. *Contrib. Mineral. Petrol.* **150**, 561–580 doi:10.1007/s00410-005-0025-8 (2005).
52. Amelin, Y., Lee, D.-G., Halliday, A. N. & Pidgeon, R. T. Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons. *Nature* **399**, 252–255 (1999).
53. Bodet, F. & Scharer, U. Evolution of the SE-Asian continent from U-Pb and Hf isotopes in single grains of zircon and baddeleyite from large rivers. *Geochim. Cosmochim. Acta* **64**, 2067–2091 (2000).
54. Vervoort, J. D., Patchett, P. J., Gehrels, G. E. & Nutman, A. P. Constraints on early Earth differentiation from hafnium and neodymium isotopes. *Nature* **379**, 624–627 (1996).
55. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
56. Harrison, T. M. *et al.* Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* **310**, 1947–1950 (2005).
57. Ireland, T. R. & Williams, I. S. in *Zircon* (eds Hancher, J. M. & Hoskin, P. W. O.) 215–238 (Reviews in Mineralogy and Geochemistry Vol. 53, Mineralogical Society of America, Washington DC, 2003).
58. Valley, J. W., Chiarenzelli, J. R. & McLelland, J. M. Oxygen isotope geochemistry of zircon. *Earth Planet. Sci. Lett.* **126**, 187–206 (1994).
59. Blichert-Toft, J., Albarède, F., Rosing, M., Frei, R. & Bridgwater, D. The Nd and Hf isotopic evolution of the mantle through the Archean. Results from the Isua supracrustals, West Greenland, and from the Birimian terranes of West Africa. *Geochim. Cosmochim. Acta* **63**, 3901–3914 (1999).
60. Moorbath, S., Whitehouse, M. J. & Kamber, B. S. Extreme Nd-isotope heterogeneity in the early Archean — fact or fiction? *Chem. Geol.* **135**, 213–231 (1997).
61. Patchett, J. P., Kouvo, O., Hedge, C. E. & Tatsumoto, M. Evolution of continental crust and mantle heterogeneity: evidence from Hf isotopes. *Contrib. Mineral. Petrol.* **78**, 279–297 (1981).
62. Knudsen, T.-L., Griffin, W. L., Hartz, E. H., Andresen, A. & Jackson, S. E. In situ hafnium and lead isotope analysis of detrital zircons from the Devonian sedimentary basin of NE Greenland: a record of repeated crustal reworking. *Contrib. Mineral. Petrol.* **141**, 83–94 (2001).
63. Griffin, W. L., Belousova, E. A., Shee, S. R., Pearson, N. J. & O'Reilly, S. Y. Archean crustal evolution in the northern Yilgarn Craton: U–Pb and Hf-isotope evidence from detrital zircons. *Precamb. Res.* **131**, 231–282 (2004).
64. Condie, K. C., Beyer, E., Belousova, E., Griffin, W. L. & O'Reilly, S. Y. U–Pb isotopic ages and Hf isotopic composition of single zircons: the search for juvenile Precambrian continental crust. *Precamb. Res.* **139**, 42–100 (2005).
65. Hawkesworth, C. J. & Kemp, A. I. S. Using hafnium and oxygen isotopes in zircons to unravel the record of crustal evolution. *Chem. Geol.* **226**, 144–162 (2006).
66. Kemp, A. I. S., Hawkesworth, C. J., Paterson, B. A. & Kinny, P. D. Episodic growth of the Gondwana supercontinent from hafnium and oxygen isotope ratios. *Nature* **439**, 580–583 (2006).
67. Rudnick, R. L. Growing from below. *Nature* **347**, 711–712 (1990).
68. Reymer, A. & Schubert, G. Phanerozoic addition rates to the continental crust and crustal growth. *Tectonics* **3**, 63–77 (1984).
69. Clift, P. & Vannuchi, P. Controls on tectonic accretion versus erosion in subduction zones: Implications for the origin and recycling of the continental crust. *Rev. Geophys.* **42**, 1–31 (2004).
70. von Huene, R. & Scholl, D. W. Observations at convergent margin concerning sediment subduction, subduction erosion and the growth of the continental crust. *Rev. Geophys.* **29**, 279–316 (1991).
71. Davies, G. F. Punctuated tectonic evolution of the Earth. *Earth Planet. Sci. Lett.* **136**, 363–379 (1995).
72. Bowring, S. A. & Housh, T. The Earth's early evolution. *Science* **269**, 1535–1540 (1995).
73. Vervoort, J. D. & Blichert-Toft, J. Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochim. Cosmochim. Acta* **63**, 533–556 (1999).
74. Storey, C. D., Jeffries, T. E. & Smith, M. Common lead-corrected laser ablation ICP-MS U–Pb systematics and geochronology of titanite. *Chem. Geol.* **227**, 37–52 (2006).
75. Elliott, T. E., Jeffcoate, A. & Bouman, C. The terrestrial Li isotope cycle: light weight constraints on mantle convection. *Earth Planet. Sci. Lett.* **220**, 231–245 (2004).
76. Jaupart, C. & Mareschal, J.-C. in *The Crust* (ed. Rudnick, R. L.) 65–84 (Treatise in Geochemistry, Vol. 3, Elsevier, Amsterdam, 2003).
77. Collerson, K. D. & Kamber, B. S. Evolution of the continents and the atmosphere. *Science* **283**, 1519–1522 (1999).
78. Taylor, S. R. & McLennan, S. M. The geochemical evolution of the continental crust. *Rev. Geophys.* **33**, 241–265 (1995).

Acknowledgements We thank J. Blundy, T. Elliott, K. Gallagher, G. Helffrich, J. Kramers and S. Turner for many discussions on these and related topics, and for reading early versions of the manuscript. R. Arculus and J. Vervoort provided comments and suggestions that greatly improved the paper. The analytical work relied heavily on the enthusiasm and expertise of the Edinburgh Ion Microprobe Facility, and of C. Coath and G. Foster in Bristol.

Author Contributions Both authors contributed equally to this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to C.J.H. (c.j.hawkesworth@bristol.ac.uk).

ARTICLES

Reconstructing the early evolution of Fungi using a six-gene phylogeny

Timothy Y. James¹, Frank Kauff¹, Conrad L. Schoch^{2*}, P. Brandon Matheny^{3*}, Valérie Hofstetter^{1*}, Cymon J. Cox^{1†}, Gail Celio⁴, Cécile Gueidan¹, Emily Fraker¹, Jolanta Miadlikowska¹, H. Thorsten Lumbsch⁵, Alexandra Rauhut⁶, Valérie Reeb¹, A. Elizabeth Arnold^{1†}, Anja Amtoft⁷, Jason E. Stajich⁸, Kentaro Hosaka^{2†}, Gi-Ho Sung², Desiree Johnson², Ben O'Rourke², Michael Crockett², Manfred Binder³, Judd M. Curtis³, Jason C. Slot³, Zheng Wang^{3†}, Andrew W. Wilson³, Arthur Schüßler⁹, Joyce E. Longcore¹⁰, Kerry O'Donnell¹¹, Sharon Mozley-Standridge¹², David Porter¹², Peter M. Letcher¹³, Martha J. Powell¹³, John W. Taylor¹⁴, Merlin M. White¹⁵, Gareth W. Griffith¹⁶, David R. Davies¹⁷, Richard A. Humber¹⁸, Joseph B. Morton¹⁹, Junta Sugiyama²⁰, Amy Y. Rossman²¹, Jack D. Rogers²², Don H. Pfister²³, David Hewitt²³, Karen Hansen²³, Sarah Hambleton²⁴, Robert A. Shoemaker²⁴, Jan Kohlmeyer²⁵, Brigitte Volkmann-Kohlmeyer²⁵, Robert A. Spotts²⁶, Maryna Serdani²⁶, Pedro W. Crous²⁷, Karen W. Hughes²⁸, Kenji Matsuura²⁹, Ewald Langer³⁰, Gitta Langer³⁰, Wendy A. Untereiner³¹, Robert Lücking⁵, Burkhard Budel⁶, David M. Geiser³², André Aptroot³³, Paul Diederich³⁴, Imke Schmitt^{5†}, Matthias Schultz³⁵, Rebecca Yahr^{1†}, David S. Hibbett³, François Lutzoni¹, David J. McLaughlin⁴, Joseph W. Spatafora² & Rytas Vilgalys¹

The ancestors of fungi are believed to be simple aquatic forms with flagellated spores, similar to members of the extant phylum Chytridiomycota (chytrids). Current classifications assume that chytrids form an early-diverging clade within the kingdom Fungi and imply a single loss of the spore flagellum, leading to the diversification of terrestrial fungi. Here we develop phylogenetic hypotheses for Fungi using data from six gene regions and nearly 200 species. Our results indicate that there may have been at least four independent losses of the flagellum in the kingdom Fungi. These losses of swimming spores coincided with the evolution of new mechanisms of spore dispersal, such as aerial dispersal in mycelial groups and polar tube eversion in the microsporidia (unicellular forms that lack mitochondria). The enigmatic microsporidia seem to be derived from an endoparasitic chytrid ancestor similar to *Rozella allomyces*, on the earliest diverging branch of the fungal phylogenetic tree.

Fungi, Viridiplantae and Animalia are all large clades descended from unicellular, flagellated, aquatic forms that radiated extensively on land. For both plants and animals, biologists have developed unified hypotheses regarding the evolution of morphology and ecology from ancestral to highly derived traits. For example, among green plants, morphologically simple photosynthetic forms, such as unicellular

green algae, gave rise to multicellular forms such as bryophytes, and were followed by a radiation of complex flowering forms with highly derived sexual mechanisms at the tips of the plant phylogeny^{1,2}. Similarly, animals seem to have evolved increasingly complex tissue systems and development from a simple, flagellated, protist-like ancestor similar to extant Choanoflagellida³.

¹Department of Biology, Duke University, Durham, North Carolina 27708-0338, USA. ²Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331-2902, USA. ³Department of Biology, Clark University, Worcester, Massachusetts 01610, USA. ⁴Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108, USA. ⁵Field Museum of Natural History, Chicago, Illinois 60605-2496, USA. ⁶Fachbereich Biologie, Abteilung Pflanzenökologie und Systematik, 67653 Kaiserslautern, Germany. ⁷Institute of Systematic Botany, New York Botanical Garden, Bronx, New York 10458-6126, USA. ⁸University Program in Genetics and Genomics, Duke University, Durham, North Carolina 27708-0338, USA. ⁹Institute of Botany, Darmstadt University of Technology, D-64287 Darmstadt, Germany. ¹⁰Department of Biological Sciences, University of Maine, Orono, Maine 04469, USA. ¹¹National Center for Agricultural Utilization Research, USDA Agricultural Research Service, Peoria, Illinois 61604, USA. ¹²Department of Plant Biology, University of Georgia, Athens, Georgia 30605, USA. ¹³Department of Biological Sciences, University of Alabama, Tuscaloosa, Alabama 35487, USA. ¹⁴Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. ¹⁵Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045-7534, USA. ¹⁶Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DA, UK. ¹⁷Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth, Ceredigion SY23 3EB, UK. ¹⁸United States Plant, Soil and Nutrition Laboratory, USDA-ARS Plant Protection Research Unit, Ithaca, New York 14853-2901, USA. ¹⁹Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia 26506-6057, USA. ²⁰TechnoSuruga, Chiyoda-ku, Tokyo 101-0052, Japan. ²¹Systematic Botany and Mycology Laboratory, USDA Agricultural Research Service, Beltsville, Maryland 20705, USA. ²²Department of Plant Pathology, Washington State University, Pullman, Washington 99164, USA. ²³Harvard University Herbaria, Cambridge, Massachusetts 02138, USA. ²⁴Biodiversity (Mycology and Botany), Agriculture and Agri-Food Canada, Ottawa, Ontario K1A 0C6, Canada. ²⁵Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, North Carolina 28557, USA. ²⁶Mid-Columbia Agricultural Research and Extension Center, Oregon State University, Hood River, Oregon 97031, USA. ²⁷Centraalbureau voor Schimmelfcultures, Fungal Biodiversity Centre, 3508 AD Utrecht, The Netherlands. ²⁸Botany Department, University of Tennessee, Knoxville, Tennessee 37996, USA. ²⁹Faculty of Agriculture, Okayama University, Okayama 700-8530, Japan. ³⁰Institut für Biologie, Universität Kassel, D-34132 Kassel, Germany. ³¹Department of Botany, Brandon University, Brandon, Manitoba R7A 6A9, Canada. ³²Department of Plant Pathology, Penn State University, University Park, Pennsylvania 16802, USA. ³³Adviesbureau voor Bryologie en Lichenologie, NL-3762 XK Soest, The Netherlands. ³⁴Musée national d'histoire naturelle, L-2160 Luxembourg. ³⁵Biozentrum Klein Flottbek und Botanischer Garten, Universität Hamburg, Systematik der Pflanzen, D-22609 Hamburg, Germany. †Present addresses: Biometry and Molecular Research, Department of Zoology, Natural History Museum, London SW7 5BD, UK (C.J.C.); Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA (A.E.A.); Department of Botany, The Field Museum, Chicago, Illinois 60605-2496, USA (K.H.); Department of Biological Sciences, Roy J. Carver Center for Comparative Genomics, University of Iowa, Iowa City, Iowa 52242, USA (Z.W.); Leibniz Institute for Natural Product Research and Infection Biology, Hans-Knöll-Institute, D-07745 Jena, Germany (I.S.); Royal Botanic Garden Edinburgh, Edinburgh EH3 5LA, UK (R.Y.).

*These authors contributed equally to this work.

Currently, no accepted phylogenetic hypothesis exists for the evolution of form and nutritional mode for the earliest fungi. Traditional views of fungal phylogeny indicate that fungi with flagellated cells (Chytridiomycota) are the sister group of the remaining phyla of non-flagellated fungi (Zygomycota, Glomeromycota, Ascomycota and Basidiomycota), implying a single loss of the flagellum coincident with a shift to land. Key adaptations to the terrestrial habit in the fungi include the evolution of a filamentous growth form and the development of aerially dispersed spores. However, recent phylogenetic studies question the monophyly of the basal phyla Chytridiomycota and Zygomycota^{4,5}. Resolving the phylogeny of the basal groups of the Fungi and their relationships to Ascomycota and Basidiomycota is necessary to understand the sequence of events leading to the colonization of land and the evolution of terrestrial ecosystems. Here we present a multilocus phylogeny of the kingdom Fungi, including representatives of all currently recognized phyla. This analysis provides a robust kingdom-level phylogeny and suggests that there were at least four independent losses of flagella during the early evolution of the Fungi.

We estimated the phylogeny of the Fungi using data from six gene regions: 18S rRNA, 28S rRNA, 5.8S rRNA, elongation factor 1- α (*EF1 α*), and two RNA polymerase II subunits (*RPB1* and *RPB2*). Incongruence among gene regions was tested by maximum likelihood bootstrap (MLBS) analyses of each data partition. This strategy allowed us to identify potential contaminant sequences in addition to conflicting phylogenetic signal. Very little conflicting signal among genes was detected, allowing construction of one super-matrix combining the data for all six gene regions for 199 fungal taxa, 29 of which used data from genome sequencing projects (Supplementary Notes 1). Only 6% of the cells in the super-matrix were missing data, and the number of aligned nucleotides was 6,436. The data were analysed by bayesian methods using a heterogeneous amino-acid and nucleotide model (see Supplementary Notes 2 for a nucleotide-only analysis). Support was estimated at nodes by bayesian posterior probabilities (BPP), MLBS and analysis of individual gene partitions (Supplementary Notes 3).

Chytridiomycota is not monophyletic

The combined gene phylogeny of the Fungi supported monophyly of the Ascomycota, Basidiomycota and Glomeromycota (Fig. 1). The Ascomycota and Basidiomycota formed a clade of 'dikarya' (that is, fungi characterized by having a portion of their life cycle with paired nuclei). Phylogenetic analyses also supported, by BPP, a clade uniting the dikarya and Glomeromycota, in agreement with previously published 18S rRNA phylogenies^{6,7}. The opisthokont clade (Fungi, Metazoa and Choanoflagellida) was also recovered, as has been reported in other studies^{3,8,9}. Two unexpected results were the placements of the endoparasitic, spizellomycetalean chytrids *Olpidium brassicae* and *R. allomyces*. *Olpidium brassicae* grouped with the Zygomycota as sister taxon to *Basidiobolus ranarum*, and *R. allomyces* grouped with the microsporidia as the earliest diverging branch of the Fungi.

The phylum Chytridiomycota consists of true fungi that produce flagellated spores (zoospores). On the basis of ultrastructural studies, the chytrid zoospore is homologous to that of non-fungal opisthokonts¹⁰. The ultrastructural complexity of the opisthokont zoospore suggests that it has evolved only once. Because the zoospore is an ancestral trait, Chytridiomycota is solely defined on a shared ancestral trait (symplesiomorphy) rather than a shared derived trait (synapomorphy). Our phylogeny indicates that the Chytridiomycota is polyphyletic (Fig. 1), consisting of early diverging lineages that have retained the zoospore. However, one large clade of Chytridiomycota uniting the orders Chytridiales, Monoblepharidales, Neocallimastigales and some Spizellomycetales (which we call the 'euchytrids') is recovered with high support values in the combined analysis as well as in multiple, single-gene-based analyses (Fig. 1 and Supplementary Notes 3).

In the present phylogeny (Fig. 1), six losses of the flagellum are inferred to have occurred during the evolution of the Fungi. Ancestral state reconstruction of the presence or absence of the flagellum along the phylogeny for each of the 58,611 credible trees demonstrated 4–6 losses (mean 5.86) of the flagellum within the Fungi. One well-supported loss took place along the branch leading to *Hyaloraphidium curvatum*, a unique fungus that grows superficially like a unicellular planktonic alga¹¹. A second loss occurred in the lineage leading to the microsporidia and 2–4 losses occurred among Zygomycota. Variation in the number of losses of the flagellum is attributable, in part, to the uncertain placement of *O. brassicae* and members of the microsporidia. Rearrangement of the phylogenetic position of *O. brassicae* and microsporidia can create phylogenies requiring only two or three losses of the flagellum; however, each of these alternative phylogenies is rejected as statistically worse (in likelihood; $P < 0.05$) than that shown in Fig. 1.

Most molecular phylogenies of the Fungi based on 18S rDNA have placed the zygomycete *Basidiobolus* among Chytridiomycota^{4,12}. This placement indicated that *Basidiobolus* might have made the transition recently from a zoospore state, and that an independent loss of a flagellum occurred in this lineage¹². This argument was strengthened by the presence in two *Basidiobolus* species of a ring-shaped spindle pole body that contains 11–12 singlet microtubules similar to a centriole, but lacks centriolar ninefold symmetry¹³. Our phylogeny is the first to place *Basidiobolus* close to Entomophthorales, the order within which it has been classified traditionally and to which it is ecologically and morphologically allied¹⁴ (for additional phylogenetic support from a paralogous copy of *EF1 α* , see Supplementary Notes 4). Unexpectedly, the phylogeny also suggests a relationship between *B. ranarum* and the chytrid *O. brassicae* (Fig. 1). A functional link between the two taxa is unclear: *O. brassicae* is an endoparasite of plant roots, whereas *Basidiobolus* is associated with insects, soil and amphibians.

Phylogenetic position of the microsporidia

Microsporidia are obligately endoparasitic, protist-like organisms with highly reduced morphology and genomes¹⁵. A defining characteristic of these parasites is the elaborate mechanism by which the spore contents are rapidly injected into the host's cytoplasm through a thin polar tube. Placement of microsporidia in the tree of life has been problematic owing to their extremely accelerated rate of sequence evolution. The earliest phylogenetic analyses of 18S rRNA placed the microsporidia among the earliest diverging lineages of eukaryotes¹⁵; however, these analyses now seem to have been an artefact of 'long branch attraction' of microsporidia to the base of the phylogeny¹⁵. More recent results using *RPB1*, α - and β -tubulin, and other genes, have suggested a fungal origin of the microsporidia^{16–18}, a placement consistent with their having the shared traits of closed mitosis and spores that contain chitin and trehalose¹⁹. Only one study has placed the microsporidia with a specific fungal lineage, in which a relationship was demonstrated between members of the Zygomycota and microsporidia by using tubulin proteins¹⁸. However, tubulin proteins seem to have evolved at different rates in flagellated and non-flagellated fungi^{18,20}.

The microsporidia and *R. allomyces* are intracellular parasites of primarily animals and fungi, respectively. A similarity between microsporidia and *R. allomyces* is the absence of a cell wall when invading host cells, such that the plasma membrane of the parasite makes direct contact with the cytoplasm of the host cell^{19,21}. Although *R. allomyces* does not seem to occupy a long phylogenetic branch, we tested whether the placement of microsporidia with *R. allomyces* was due to long branch attraction. Two different methods suggested that the relationship between microsporidia and *R. allomyces* is not due to long branch attraction (see Supplementary Notes 5). We also tested whether alternative placements for the microsporidia could be statistically rejected from the maximum likelihood phylogeny shown in Fig. 1 using the approximately unbiased test²². Alternative placements of microsporidia with Fungi that have been suggested

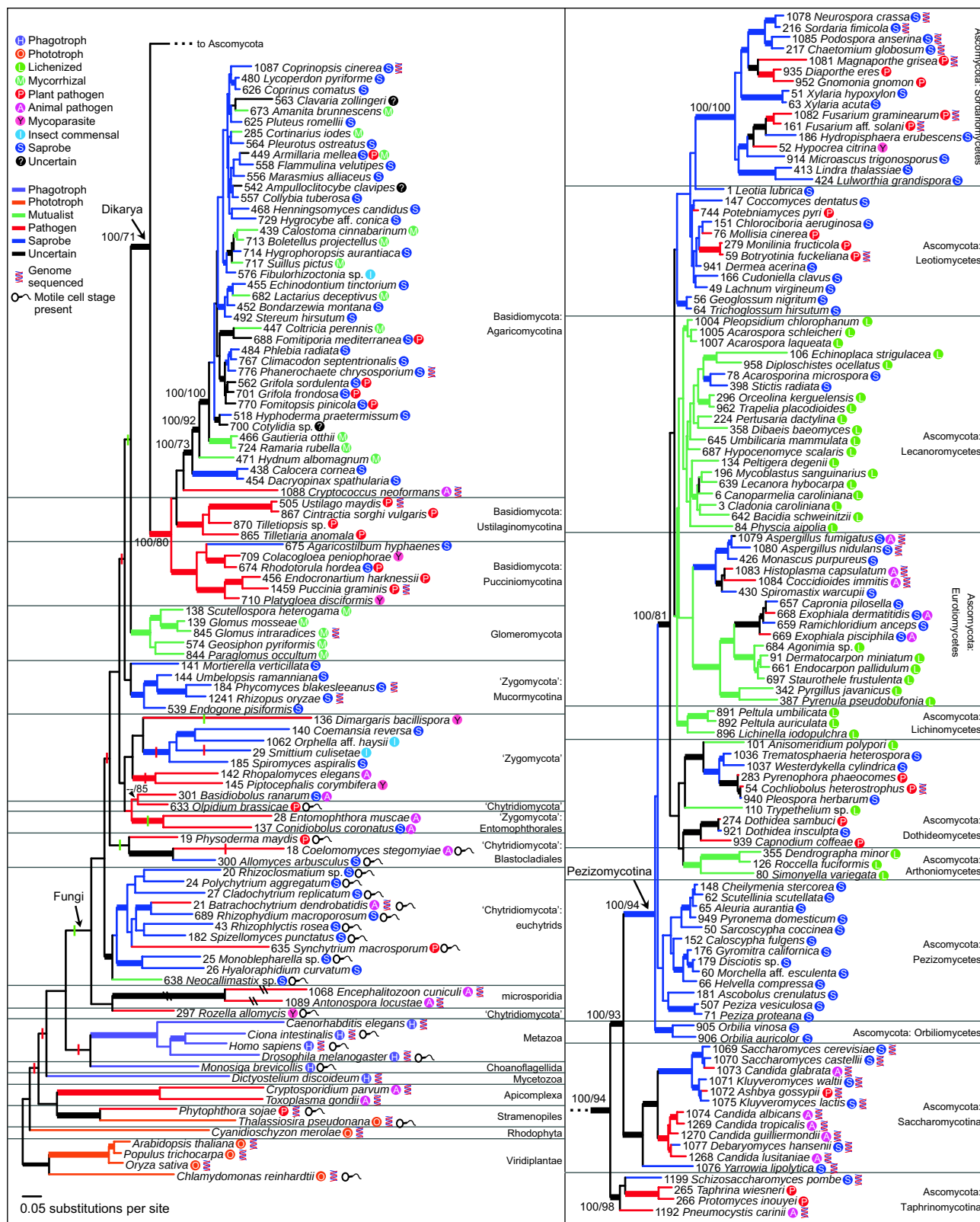


Figure 1 | Phylogeny of the kingdom Fungi using bayesian analysis of the combined, six-gene data set. Each fungal species begins with a unique 'Assembling the Fungal Tree of Life' identifier, followed by genus and species. Indicated for each terminal taxon are: nutritional mode, whether they produce flagellated cells and if there is a genome sequence for the taxon completed or underway. Thickened branches indicate those that are supported both by heterogeneous bayesian analysis (BPP $\geq 95\%$) and by MLBS ($\geq 70\%$). Almost every branch was supported by BPP and thus values are not shown. Where indicated, support values (percentage of trees in

agreement out of 58,611 trees) indicate BPP followed by MLBS. Branches are shaded according to reconstruction of nutritional mode. Microsporidia branches have been shortened three times (double black break) to increase readability. Red vertical ticks on branches indicate alternative placements of microsporidia that might be significantly rejected ($P < 0.05$) and green ticks indicate placements that cannot be rejected. Quotation marks indicate non-monophyly of the taxon. The name 'Mucormycotina' will be validated in a manuscript that is in preparation.

include: a sister relationship to the dikarya²³; sister to the zygomycete order Entomophthorales¹⁸; and among the harpellid Trichomycetes¹⁹, represented here by *Smittium culisetae*. We were able to reject ($P < 0.05$) nine alternative placements of the microsporidia (red vertical ticks in Fig. 1), including early divergences among eukaryotes. However, we were unable to reject a placement of microsporidia as sister to Entomophthorales, as sister to the blastocladiacean chytrids, as sister to the zygomycete *Dimargaris*, as sister to dikarya and as sister to the Fungi (green vertical ticks in Fig. 1).

Taken together, our results suggest that the relationship between the microsporidia and *R. allomyis* is a result of true phylogenetic signal. The present phylogeny provides an alternative hypothesis for the placement of microsporidia, specifically on the earliest diverging fungal branch with the chytrid *R. allomyis*. However, support for this relationship is derived only from the *RPB1* and *RPB2* gene partitions and is not supported by rDNA (see Supplementary Notes 3); alternative hypotheses in which the microsporidia diverge among early fungi cannot be rejected. The ultimate resolution of the placement of microsporidia will require sampling of additional genes from basal fungal taxa.

Dikarya

The majority (~98%) of described fungal species are members of the dikarya clade, which includes the two phyla Ascomycota and Basidiomycota. Ascomycota is the largest phylum within the Fungi and is characterized by the production of meiospores (ascospores) in specialized sac-shaped meiosporangia (asci), which may or may not be produced within a sporocarp (ascoma). Ascomycota is divided into three monophyletic subphyla: Taphrinomycotina, Saccharomycotina and Pezizomycotina (each of which is well supported as monophyletic in the phylogeny; Fig. 1). Taphrinomycotina is resolved as the earliest diverging clade; it includes a diverse group of species that exhibit yeast-like (for example, *Pneumocystis*) and dimorphic—that is, yeast-like and filamentous (for example, *Taphrina*)—growth forms. The subphylum Saccharomycotina consists of the ‘true yeasts’, including bakers’ yeast (*Saccharomyces cerevisiae*) and *Candida albicans*, the most frequently encountered fungal pathogen of humans. Pezizomycotina is the largest subphylum of Ascomycota and includes the vast majority of filamentous, fruit-body-producing species. Data presented here resolved the Orbiliomycetes and Pezizomycetes as the early-diverging lineages of the Pezizomycotina, with the remaining seven classes sampled forming a well-supported crown clade. Reduced ascomatal morphologies, whereby asci are contained within fruit bodies that are enclosed partially (Dothideomycetes, Eurotiomycetes and some Sordariomycetes) or completely (Eurotiomycetes, Leotiomycetes and some Sordariomycetes), are restricted to the crown clade of Pezizomycotina.

The Basidiomycota includes about 30,000 species of rusts, smuts, yeasts, and mushroom fungi²⁴. Most are characterized by meiospores (basidiospores) on the exterior of typically club-shaped meiosporangia (basidia). Phylogenetic relationships among the three subphyla of Basidiomycota are uncertain. The subphylum Pucciniomycotina is primarily distinguished by containing the rust fungi (7,000 species), which are primarily pathogens of land plants. Cytological and biochemical data²⁵ are consistent with a sister group relationship between the subphyla Ustilaginomycotina and Agaricomycotina, as shown in Fig. 1. The Ustilaginomycotina includes 1,500 species of true smut fungi and yeasts, most of which cause systemic infections of angiosperm hosts. The Agaricomycotina includes almost two-thirds of known basidiomycetes, including the vast majority of mushroom-forming fungi. Much of the morphological diversity exemplified in mushroom fruiting bodies is the result of radiations of certain lineages within the Agaricomycotina, and recovering their relationships with confidence has proven difficult^{26,27}. Early-diverging lineages in the Agaricomycotina, which are strongly supported in Fig. 1, also include parasitic and/or saprotrophic fungi capable of dimorphism or yeast-like phases. The mycorrhizal basidiomycetes

seem to have multiple, independent evolutionary origins from saprotrophic ancestors as previously suggested²⁸.

Characteristics of early fungi

We reconstructed ancestral states for major nutritional modes in the Fungi using maximum likelihood (Fig. 1). Most of the ancestral character states of deep nodes are equivocal, with the exception of the common ancestor of members of the Basidiomycota, for which a parasitic ancestor is suggested. The phylogeny suggests that numerous transitions from a pathogenic to a saprophytic nutritional mode have occurred, as well as the reverse (Fig. 1). Although the nutritional mode of the common ancestor of Fungi is ambiguous, the earliest diverging branch in the Fungi contains parasitic species (*R. allomyis* and microsporidia). Recent studies^{9,29} showed that the closest known relative to Fungi is the amoeboid protist *Nuclearia*, which grows phagotrophically on algae and bacteria. Amoeboid phases are also observed in basal fungi: *Rozella* seems to phagocytose the organelles of its host³⁰ and many chytrid zoospores undergo an amoeboid, motile phase before encysting. After the divergence of the *Rozella* and microsporidia lineage, the remaining fungi evolved filamentous growth (for example, hyphae and rhizoids), which aids in substrate attachment and absorptive nutrition involving extracellular digestion. Within the Basidiomycota and Ascomycota, a reversion to a unicellular, yeast-like growth form is observed among the earliest diverging lineages, perhaps implicating a prior advantage for this growth form in the early history of the Fungi.

It is unclear whether the common ancestor of Fungi was marine. Most zoosporic true fungi, including all of the chytrids sampled in this study, grow in freshwater or soil habitats. Therefore, the diversification of the major lineages (phyla) within the kingdom Fungi probably occurred in a terrestrial environment but before the emergence of land plants^{31,32}. Mycorrhiza-like symbioses of the phylum Glomeromycota are suggested to have been crucial in the colonization of land by plants³³. Extant members of the Glomeromycota live exclusively as obligate symbionts of photoautotrophs, including not only vascular plants and bryophytes, but also cyanobacteria. This raises the hypothesis that terrestrial members of the Glomeromycota living symbiotically with cyanobacteria or algae, in semi-aquatic and humid habitats later became the symbiotic partners of early land plants³⁴.

The present multilocus phylogeny explains the possible morphology and ecology of early fungi. The early-diverging lineages consist of a grade of zoosporic fungi, suggesting that the earliest fungi were primarily aquatic and lacked aerial spore dispersal. The loss of flagellated spores is inferred to have occurred at least four times. Each loss seems to have coincided with novel innovations in spore production and dispersal: microscopic wind-dispersed spores in terrestrial fungi; forcibly discharged conidia in the Entomophthorales; non-flagellated, mitotically produced spores in the planktonic *Hyaloraphidium curvatum*; and a complex polar tube apparatus in microsporidia. The sister kingdom to the Fungi (Animalia) evolved diverse body plans capable of feeding by ingestion, whereas the fungal branch developed a myriad of unicellular and filamentous forms optimized for absorptive nutrition. With a well-resolved phylogeny, fungal biologists can now study the evolution of complexity and multicellularity, and compare the evolution of these traits in fungi with their evolution in plants and animals.

METHODS

Molecular techniques. Sequence data were generated from 170 fungal species, primarily using pure cultures and herbarium material (Supplementary Notes 1). We used standard polymerase chain reaction (PCR) protocols²⁵ for amplification and sequencing of six gene regions: the 18S ribosomal RNA gene (nearly full length), the 28S ribosomal RNA gene (primers LR0R and LR7), the internal transcribed spacer (ITS) RNA gene region (full length), *EF1 α* (mostly primers EF1-983F and EF1-2218R), RNA polymerase II largest subunit (*RPB1*, mostly primers RPB1-Af and RPB1-G2R) and RNA polymerase II second largest subunit (*RPB2*, primers RPB2-5F and RPB2-11bR). Information on the PCR primers can be found at <http://www.aftol.org/primers.php>. In a number of basal

fungal taxa, the *EF1 α* gene was not detected, but a paralogous copy of the gene was recovered (*EFL*, or the *EF1 α* -like gene³⁵; see Supplementary Notes 4). We also obtained sequences from fungal and eukaryotic genomes by retrieving sequences from GenBank and genome servers. Although our data set contains both partial sequences and missing data points, in the case of only one taxon (the choanoflagellate *Monosiga brevicollis*) were fewer than four genes sampled.

Phylogenetic reconstruction. The data set consisted of 214 taxa, 199 of which were fungi. Sequences were aligned and ambiguous regions excluded in MacClade³⁶. Conflict among the six genes was assessed by separate MLBS of each data partition using 250 bootstrap replicates in PHYML³⁷. We ignored two conflicts, one including microsporidian 18S sequences (known to be subject to long branch attraction) and the other involving marginally conflicting signal of the Pyrenulales (Ascomycota). Data were combined into one matrix with *EF1 α* , *RPB1* and *RPB2* translated into amino acids and 18S, 28S and 5.8S as nucleotides. We applied a heterogeneous maximum likelihood model to the data set with six unlinked partitions, one for each gene. The 18S and 28S genes were fitted to a general-time-reversible model with a proportion of invariant sites and gamma distributed rates (GTR+I+ Γ), the 5.8S data used GTR+ Γ and proteins used the JTT+I+ Γ fixed rate model. The gamma distribution was approximated using four rate classes. We used MrBayes 3.1.1 (ref. 38) for phylogenetic estimation. Five independent runs were conducted (each with four chains) for 9.5×10^6 generations, sampling every 500 generations. Runs were discarded if they failed to reach the same likelihood plateau observed in other independent runs. We computed the consensus of the sampled trees, the posterior probabilities of clades, and average branch lengths from runs that converged to the same likelihood plateau (58,611 trees). For the analysis of the combined super-matrix we also tested for convergence of runs by analysing frequencies of splits using the software AWTY³⁹ and found that the consensus topology constructed using this criterion trivially differed from that based on log likelihood scores. We also assessed support for nodes on the nucleotide data (third codon positions excluded) by MLBS (500 replicates) using PHYML with a GTR+I+ Γ model. Tests for statistical differences in likelihoods of alternative topologies were assessed using the approximately unbiased test²² on the nucleotide data with site-wise, log-likelihood values calculated using TREE-PUZZLE v5.2 (ref. 40).

Ancestral character state reconstruction of nutritional mode was conducted using the maximum likelihood model Mk1 in Mesquite 1.0 (ref. 41). Taxa were assigned to ecological character states on the basis of published literature, resolving ambiguous assignments when possible. Reconstructions are reported for only those branches significantly assigned an unequivocal character state in a majority of 1,000 trees randomly drawn from the sample of credible trees. The number of losses of the flagellum within the Fungi was also estimated for all 58,611 credible trees using Dollo parsimony as implemented in MacClade.

Received 4 May; accepted 25 July 2006.

- Karol, K. G., McCourt, R. M., Cimino, M. T. & Delwiche, C. F. The closest living relatives of land plants. *Science* **294**, 2351–2353 (2001).
- Groth-Malonek, M., Pruchner, D., Grewe, F. & Knoop, V. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol. Biol. Evol.* **22**, 117–125 (2005).
- Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12**, 1773–1778 (2002).
- James, T. Y., Porter, D., Leander, C. A., Vilgaly, R. & Longcore, J. E. Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Can. J. Bot.* **78**, 336–350 (2000).
- Tanabe, Y., Saikawa, M., Watanabe, M. M. & Sugiyama, J. Molecular phylogeny of Zygomycota based on EF-1 α and RPB1 sequences: limitations and utility of alternative markers to rDNA. *Mol. Phylogenet. Evol.* **30**, 438–449 (2004).
- Schüßler, A., Schwarzott, D. & Walker, C. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol. Res.* **105**, 1413–1421 (2001).
- Tehler, A., Little, D. P. & Farris, J. S. The full-length phylogenetic tree from 1551 ribosomal sequences of chitinous fungi. *Fungi. Mycol. Res.* **107**, 901–916 (2003).
- Cavalier-Smith, T. & Chao, E. E. The opalozoan *Apusomonas* is related to the common ancestor of animals, fungi, and choanoflagellates. *Proc. R. Soc. Lond. B* **261**, 1–9 (1995).
- Steenkamp, E. T., Wright, J. & Baldauf, S. L. The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106 (2005).
- Barr, D. J. S. Evolution and kingdoms of organisms from the perspective of a mycologist. *Mycologia* **84**, 1–11 (1992).
- Ustinova, I., Krienitz, L. & Huss, V. A. R. *Hyaloraphidium curvatum* is not a green alga, but a lower fungus; *Amoebidium parasiticum* is not a fungus, but a member of the DRIPs. *Protist* **151**, 253–262 (2000).
- Nagahama, T., Sato, H., Shimazu, M. & Sugiyama, J. Phylogenetic divergence of the entomophthoralean fungi: evidence from nuclear 18S ribosomal RNA gene sequences. *Mycologia* **87**, 203–209 (1995).
- McKerracher, L. J. & Heath, I. B. The structure and cycle of the nucleus-associated organelle in two species of *Basidiobolus*. *Mycologia* **77**, 412–417 (1985).

- Blackwell, M. & Malloch, D. Similarity of *Amphoromorpha* and secondary capillitonia of *Basidiobolus*. *Mycologia* **81**, 735–741 (1989).
- Keeling, P. J. & Fast, N. M. in *Insect–Fungal Associations: Ecology and Evolution* (eds Vega, F. E. & Blackwell, M.) 97–118 (Oxford Univ. Press, Oxford, 2005).
- Hirt, R. P. et al. Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
- Katinka, M. D. et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
- Keeling, P. J. Congruent evidence from α -tubulin and β -tubulin gene phylogenies for a zygomycete origin of microsporidia. *Fungal Genet. Biol.* **38**, 298–309 (2003).
- Cavalier-Smith, T. in *The Mycota* (eds McLaughlin, D. J., McLaughlin, E. G. & Lemke, P. A.) 3–37 (Springer, New York, 2001).
- Corradi, N., Hijiri, M., Fumagalli, L. & Sanders, I. R. Arbuscular mycorrhizal fungi (Glomeromycota) harbour ancient fungal tubulin genes that resemble those of the chytrids (Chytridiomycota). *Fungal Genet. Biol.* **41**, 1037–1045 (2004).
- Held, A. A. *Rozella* and *Rozellopsis*: Naked endoparasitic fungi which dress up as their hosts. *Bot. Rev.* **47**, 451–515 (1981).
- Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
- Gill, E. E. & Fast, N. M. Assessing the microsporidia–fungi relationship: combined phylogenetic analysis of eight genes. *Gene* **375**, 103–109 (2006).
- Kirk, P. M., Cannon, P. F., David, J. C. & Stalpers, J. A. (eds) *Ainsworth & Bisby's; Dictionary of the Fungi* 60 (CAB International, Wallingford, UK, 2001).
- Lutzoni, F. et al. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *Am. J. Bot.* **91**, 1446–1480 (2004).
- Binder, M. et al. The phylogenetic distribution of resupinate forms across the major clades of mushroom-forming fungi (Homobasidiomycetes). *Systematics and Biodiversity* **3**, 113–157 (2005).
- Moncalvo, J.-M. et al. One hundred and seventeen clades of euagarics. *Mol. Phylogenet. Evol.* **23**, 357–400 (2002).
- Hibbett, D. S., Gilbert, L.-B. & Donoghue, M. J. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* **407**, 506–508 (2000).
- Medina, M. et al. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int. J. Astrobiol.* **2**, 203–211 (2003).
- Powell, M. J. Fine structure of the unwall thallus of *Rozella polyphagi* in its host *Polyphagus euglenae*. *Mycologia* **76**, 1039–1048 (1984).
- Berbee, M. L. & Taylor, J. W. in *The Mycota* (eds McLaughlin, D. J., McLaughlin, E. G. & Lemke, P. A.) 229–245 (Springer, New York, 2001).
- Heckman, D. S. et al. Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**, 1129–1133 (2001).
- Pirozynski, K. A. & Malloch, D. W. The origin of land plants: a matter of mycotrophism. *Biosystems* **5**, 153–164 (1975).
- Schüßler, A. Molecular phylogeny, taxonomy, and evolution of arbuscular mycorrhiza fungi and *Geosiphon pyriformis*. *Plant Soil* **244**, 75–83 (2002).
- Keeling, P. J. & Inagaki, Y. A class of eukaryotic GTPase with a punctuate distribution suggesting multiple functional replacements of translation elongation factor 1 α . *Proc. Natl Acad. Sci. USA* **101**, 15380–15384 (2004).
- Maddison, D. & Maddison, W. *MacClade Version 4.05: Analysis of Phylogeny and Character Evolution* (Sinauer Associates, Sunderland, Massachusetts, USA, 2002).
- Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
- Huelsenbeck, J. P. & Ronquist, F. MrBayes: bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Wilgenbusch, J. C., Warren, D. L. & Swofford, D. L. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty> (2004).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
- Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis. Version 1.0. <http://mesquiteproject.org> (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for the project was provided by the National Science Foundation's 'Assembling the Tree of Life' and 'Research Coordination Network' programs. For technical assistance we thank L. Bukovnik, C. Roberts and H. Matthews. We also thank the following individuals for sharing research materials: M. C. Aime, W. R. Buck, M. S. Cole, P. Crane, Y. Dalpe, D. M. Hillis, S. L. Joneson, R. Petersen, C. Printzen, E. Vellinga, H. Whisler and A. Zavarzin. We are very thankful to B. Mueller, J. Harer, B. Rankin, J. Pormann and S. Dilda for providing access to the Duke CSEM computer cluster. P. Keeling provided unpublished information used to analyse EFL in Fungi.

Author Information Data for this project have been deposited in GenBank (see Supplementary Notes 1 for accession numbers), and the alignments can be accessed on the Assembling the Fungal Tree of Life website at <http://www.aftol.org/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.Y.J. (tyj2@duke.edu) or R.V. (fungi@duke.edu).

The Mg-chelatase H subunit is an abscisic acid receptor

Yuan-Yue Shen^{1*}, Xiao-Fang Wang^{1*}, Fu-Qing Wu^{1*}, Shu-Yuan Du¹, Zheng Cao¹, Yi Shang¹, Xiu-Ling Wang¹, Chang-Cao Peng¹, Xiang-Chun Yu¹, Sai-Yong Zhu¹, Ren-Chun Fan¹, Yan-Hong Xu¹ & Da-Peng Zhang¹

Abscissic acid (ABA) is a vital phytohormone that regulates mainly stomatal aperture and seed development, but ABA receptors involved in these processes have yet to be determined. We previously identified from broad bean an ABA-binding protein (ABAR) potentially involved in stomatal signalling, the gene for which encodes the H subunit of Mg-chelatase (CHLH), which is a key component in both chlorophyll biosynthesis and plastid-to-nucleus signalling. Here we show that *Arabidopsis* ABAR/CHLH specifically binds ABA, and mediates ABA signalling as a positive regulator in seed germination, post-germination growth and stomatal movement, showing that ABAR/CHLH is an ABA receptor. We show also that ABAR/CHLH is a ubiquitous protein expressed in both green and non-green tissues, indicating that it might be able to perceive the ABA signal at the whole-plant level.

The phytohormone ABA has a vital function in plant adaptation to stressful environments by regulating stomatal aperture and the expression of stress-responsive genes, and in plant development such as seed maturation, germination and seedling growth^{1–3}. Genetic approaches have permitted the characterization of numerous components involved in ABA signalling but have failed to identify ABA receptors^{1–3}. Biochemical approaches provide another way to isolate ABA receptors by the identification of ABA-binding proteins that are putative ABA receptors^{4–6}. Recently, the RNA-binding protein FCA, a homologue of an ABA-binding protein ABAP1 (ref. 6), was identified as an ABA receptor in the regulation of flowering time⁷. However, ABA receptors involved in seed development, seedling growth and stomatal movement remain elusive.

We previously reported an ABA-specific-binding protein from broad bean (*Vicia faba*) and found that this protein was potentially involved in ABA-induced stomatal signalling⁵. So we designated it ABAR (for putative abscisic acid receptor). On the basis of the sequencing information, we isolated from broad bean leaves a complementary DNA fragment encoding the carboxy-terminal half of about 770 amino acids of the putative H subunit (CHLH) of the magnesium-protoporphyrin-IX chelatase (Mg-chelatase). CHLH has been reported to have multiple functions in plant cells. In addition to its enzymatic functions as a subunit of the Mg-chelatase in producing photosynthetic apparatus⁸, CHLH has a key function in mediating plastid-to-nucleus signalling^{9–12}. We found that the yeast-expressed product of the cDNA fragment encoding the C-terminal portion of the broad bean ABAR or CHLH binds ABA specifically (data not shown). The information from *Vicia faba* led us to analyse the functions of ABAR in ABA signalling in *Arabidopsis thaliana*. Here we show that the *Arabidopsis* ABAR/CHLH (the *Arabidopsis* genomic locus tag for CHLH is At5g13630, and the GenBank accession numbers are AY070133, BT002311, NM_121366, Z68495 or AY078971) is an ABA receptor that regulates seed development, post-germination growth and stomatal aperture.

ABAR binds ABA specifically

The purified yeast-expressed *Arabidopsis* ABAR binds ABA as a saturable process (Fig. 1a). The ABAR protein possesses one binding

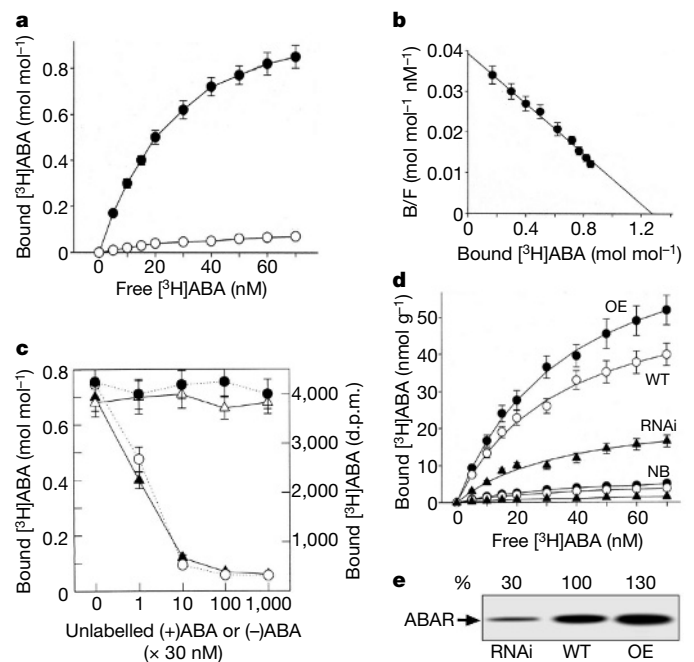


Figure 1 | ABAR binds ABA. **a**, Saturable ABA-specific binding (filled circles) to the pure yeast-expressed ABAR. Open circles, non-specific binding. **b**, Scatchard plot of binding data in **a**. B, $[^3\text{H}]\text{ABA}$ bound; F, free $[^3\text{H}]\text{ABA}$. The parameters of the curve are $K_d = 32$ nM; $B_{\text{max}} = 1.28$ mol mol⁻¹; $R^2 = 0.98$. **c**, Displacement of $[^3\text{H}](+)\text{-ABA}$ binding by (+)-ABA and (-)-ABA. Filled triangles, *in vitro* (+)-ABA binding (mol mol⁻¹); open triangles, *in vitro* (-)-ABA binding (mol mol⁻¹); filled circles, (-)-ABA pull-down assay (d.p.m.); open circles, (+)-ABA pull-down assay (d.p.m.). **d**, Saturable ABA-specific binding to the extracts of leaves of *gl1* (WT), RNAi (line 12) and overexpressing (OE, line 1) lines. NB, non-specific binding. **e**, Immunosignal of ABAR protein in the extracts in **d** and the band intensities relative to the wild type. Each point in **a–d** is the mean $\pm$ s.d. ($n = 10$).

¹China State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, 100094 Beijing, China.

*These authors contributed equally to this work.

site, as shown by a linear Scatchard plot and a maximum binding ($B_{\max}$) of 1.28 mol ABA per mol of protein, and has a high binding affinity for ABA, as shown by its equilibrium dissociation constant (K_d) of 32 nM (Fig. 1b). Furthermore, the purified ABAR binds ABA in a highly stereospecific manner, which was revealed by the efficient displacement of [3 H](+)-ABA binding by the physiologically active form (+)-ABA but not by two inactive ABA isomers, (–)-ABA and *trans*-ABA, which are structurally similar to (+)-ABA (Fig. 1c; data for *trans*-ABA not shown).

We further analysed the ABA-binding ability of natural ABAR protein. The saturable process of ABA binding to the extracts of leaves was found (Fig. 1d). Upregulation of the ABAR level by overexpressing *ABAR* enhanced, but downregulation by RNAi-mediated interference (RNAi) reduced, the $B_{\max}$ of ABA binding, whereas neither substantially changed K_d (from 35 to 38 nM) (Fig. 1d, e, and Supplementary Fig. 1), revealing that the changes in ABAR abundance alter the numbers of ABA-binding sites but do not modify the affinity. A pull-down assay with the ABAR-specific antiserum specifically co-precipitated the ABA-binding activities proportionally to the amounts of the ABAR protein (Supplementary Fig. 2), and the pulled-down ABA-binding activity was shown to be highly stereospecific for (+)-ABA (Fig. 1c; data for *trans*-ABA not shown). The data reveal that ABAR specifically binds ABA, and the binding meets primary criteria for an ABA receptor.

ABAR mediates ABA signalling as a positive regulator

To explore the functions of ABAR in ABA signalling, we generated *Arabidopsis* transgenic RNAi, antisense and overexpression lines. The plants underexpressing *ABAR* as a result of RNAi showed significant ABA-insensitive phenotypes in seed germination (Fig. 2a), post-germination growth arrest by ABA (Fig. 2b) and ABA-induced promotion of stomatal closure and inhibition of stomatal opening (Fig. 2c).

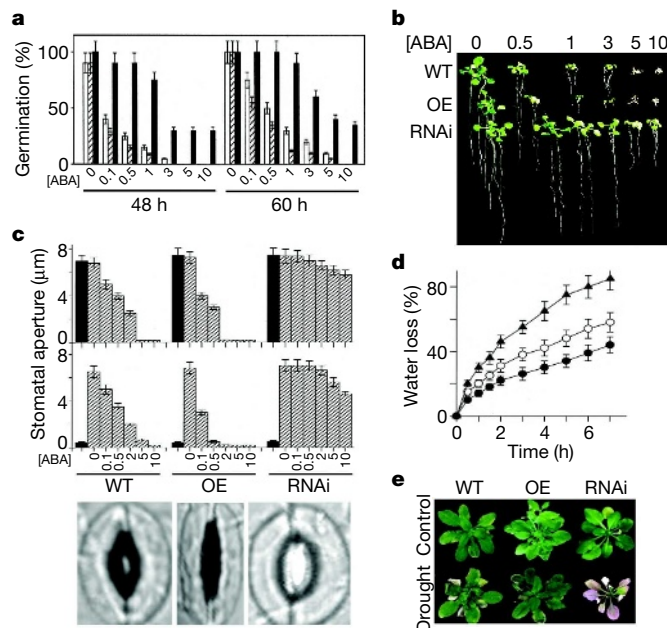


Figure 2 | Changes in *ABAR* expression alter plant sensitivity to ABA. **a, b,** Seed germination (**a**) and seedling (10 days in ABA) growth (**b**) of gl1 (WT, white columns in **a**), RNAi (black columns in **a**) and overexpressing (OE, hatched columns in **a**) lines in medium containing ($\pm$)-ABA. ABA concentrations are in μ M. **c,** ABA-induced stomatal closure (top) and inhibition of opening (middle). Black columns, initial stomatal aperture; grey columns, apertures after ABA treatment. ABA concentrations are in μ M. Bottom: stomatal apertures in the assay of inhibition of opening by 2 μ M ABA. **d,** Water loss from detached leaves. Open circles, WT; filled circles, OE; triangles, RNAi. **e,** Plant status after drought treatment (controls were fully watered plants). The RNAi line 12 and OE line 1 were used. Error bars indicate s.d. ($n = 5$).

(Fig. 2c). In contrast, the plants overexpressing *ABAR* displayed the ABA-hypersensitive phenotypes (Fig. 2a–c) and were more resistant to dehydration from their leaves or whole plants, but the RNAi plants were more sensitive to dehydration (Fig. 2d, e). Overall, the *ABAR* levels were negatively correlated with the intensity of the ABA-insensitive phenotypes (Supplementary Fig. 3). The ABA concentrations did not change in the transgenic plants (in the range of 0.2 μ g g $^{-1}$ dry weight), showing that *ABAR* is not involved in ABA biosynthesis.

We further identified a transferred DNA (T-DNA) insertion mutant in the *ABAR* gene (Supplementary Fig. 4), designated *abar-1*. Homozygous *abar-1* is lethal. The *abar-1* seeds are deficient in lipid and mature protein bodies (Supplementary Fig. 4), indicating a possible distortion of late embryonic development². These phenotypes are similar to those of the mutations in ABA-signalling genes such as *ABI3*, which has specific effects on seed maturation^{13,14}. Taken together, the data show that *ABAR* mediates ABA signalling as a positive regulator.

ABAR-mediated ABA signalling is a distinct process

Mg-chelatase, which is composed of three subunits, namely CHLD, CHLI and CHLH, catalyses the insertion of Mg²⁺ into protoporphyrin-IX (Proto) to form Mg-protoporphyrin-IX (MgProto), the first step unique to chlorophyll synthesis⁸. CHLH has a central function as a monomeric Proto-binding protein^{8,15}. The *Arabidopsis* *gun5* mutant, resulting in a single amino acid Ala 990→Val mutation in CHLH, showed that CHLH is involved in plastid-to-nucleus retrograde signalling by controlling the metabolism of the tetrapyrrole signal MgProto or by sensing the signal^{9–12}. We observed that treatment with exogenous ABA significantly decreased both the chlorophyll and Proto contents but stimulated *ABAR* expression and Mg-chelatase activity and enhanced the MgProto contents (Fig. 3a). This positive regulation of *ABAR* by ABA seems to support its function as an ABA sensor and indicates that ABA-induced chlorophyll decrease might not be attributable to the action of ABA on *ABAR*. However, as an ABA and Proto dual-ligand-binding protein, *ABAR* binds ABA independently of Proto (Supplementary Fig. 5), indicating that ABA signal perception might be distinct from Proto binding.

We observed, in a pharmaceutical assay using both norflurazon¹⁶ (an inhibitor of the carotenoid biosynthetic enzyme phytoene desaturase that causes photo-oxidative damage to chloroplasts) and chloramphenicol (CP; an inhibitor of plastid translation), that an ABA-insensitive stomatal movement occurred in parallel with a decrease in *ABAR* levels (Supplementary Fig. 7), but no correlation of ABA-responsive stomatal movement with chlorophyll or MgProto contents was found (Supplementary Figs 6, 7). We also used a chemical-regulated inducible RNAi system¹⁷ to investigate *ABAR*-mediated ABA signalling. After induction by 17 β -oestradiol, a decrease in the *ABAR* levels was observed without an alteration in the chlorophyll and MgProto contents (Fig. 3b, and data not shown), and this decrease in *ABAR* levels induced a parallel insensitivity of stomatal movement to ABA (Fig. 3b). Using the protoplasts prepared from the inducible RNAi plants, we found that a decline of *ABAR* expression repressed the mRNA levels of *RD29A* (ref. 18), *MYB2* and *MYC2* (ref. 19)—genes that respond positively to ABA—but upregulated two negative regulators of ABA signalling, namely *ABI1* (refs 20–22) and *ABI2* (refs 22, 23) (Supplementary Fig. 8). These approaches provide additional, more direct, evidence for functions of *ABAR* as a positive regulator of ABA signalling independently of chlorophyll and MgProto.

Further assays were performed with a series of mutants defective in chlorophyll metabolism or plastid signalling. *hyl* (refs 24, 25) and *hy2* (ref. 26) mutants, containing lesions in haem oxygenase and phytylcholine synthase genes, respectively, are alleles of two *gun* mutants, *gun2* and *gun3* respectively, that are defective in plastid signalling (refs 9, 10, 12). The *gun4* mutant has a lesion in a second

Proto-binding protein-encoding gene *GUN4* (refs 27, 28). *chl* mutants contain lesions in the gene encoding chlorophyll *a* oxygenase²⁹. *cch* is also a *gun* mutant and an allele of *gun5* but with a single nucleotide substitution at a different site, resulting in a single amino acid mutation Pro 642→Leu (ref. 9). These mutants have lower chlorophyll contents except *gun5*, which possesses a chlorophyll level comparable to that of its wild type (Col, Fig. 3c), and they have ABAR levels comparable to those of their wild-type plants except the *cch* mutant, which has a lower level (Fig. 3c). All the *gun* mutants *hy1/gun2*, *hy2/gun3*, *gun4*, *gun5* and *cch* have been shown to be involved in the same MgProto-triggered plastid-signalling pathway¹¹, but only the *cch* mutant had the ABA-insensitive phenotypes in germination (Fig. 3d), seedling growth (Fig. 3e) and ABA-induced stomatal movement (Fig. 3f). In all the mutants, no significant correlation of the chlorophyll levels with the ABA-responsive phenotypes was observed (Fig. 3c–f). The *cch* mutation, but not *gun5*, significantly decreased the ABA-binding activity of ABAR (Supplementary Fig. 9), which may explain the ABA-insensitive phenotypes in the *cch* mutant and wild-type phenotypes in the *gun5* mutant. Taken together, these data show clearly that ABAR is a positive regulator in ABA signal transduction involved in a signalling process that is distinct from

chlorophyll metabolism and MgProto-mediated plastid retrograde signalling.

ABAR is ubiquitous and regulates ABA-signalling genes

The *CHLH* expression was previously reported to be limited to the green tissues^{8,10}. Available data at the Genevestigator site (<http://www.genevestigator.ethz.ch>) show the presence of *Arabidopsis* *CHLH* mRNA in seeds. We found that ABAR/CHLH is a protein that is expressed ubiquitously in the non-green tissues, including the roots (Fig. 4a). ABAR might therefore function at the whole-plant level.

We found that downregulation of ABAR expression by RNAi decreased the levels of the positive regulators of ABA signalling *RD29A* (ref. 18), *MYB2* (ref. 19), *MYC2* (ref. 19), *ABI4* (refs 30, 31), *ABI5* (refs 30, 32) and *OST1* (ref. 33), but enhanced the levels of three negative regulators, *ABI1* (refs 20–22), *ABI2* (refs 22, 23) and *CIPK15* (ref. 34) in leaves (Fig. 4b). These results are essentially consistent with those from the inducible RNAi protoplasts (Supplementary Fig. 8). The seed-specific ABA-signalling genes *ABI3* (refs 13, 14), *ABI4* (refs 30, 31) and *ABI5* (refs 30, 32) and their downstream genes *EM1* and *EM6*, which are both responsible for late embryogenesis^{2,35}, were all downregulated in the siliques of the RNAi plants (Fig. 4c). In most cases the ABAR-overexpressing plants regulated these genes in a manner contrary to the RNAi plants (Fig. 4b, c). The expression levels of these genes in the *gun5* mutant were similar to those in the wild-type Columbia (data not shown). The regulation of these ABA-signalling genes by ABAR supports the contention that ABAR is a positive regulator and indicates that it might function through various pathways.

Discussion

Previous studies have shown a multiplicity and complexity of ABA perception sites that may act at the outside or inside of cells to mediate different biological functions in plants^{1–3,7}. We show that ABAR is an ABA receptor to perceive the ABA signal in seed germination, post-germination growth and stomatal movement, essentially from the following evidence: first, ABAR specifically binds ABA; second, transgenic downregulation of ABAR expression results in a decline in the number of ABA-binding sites and leads to ABA-insensitive phenotypes; third, ABAR-overexpressing plants have

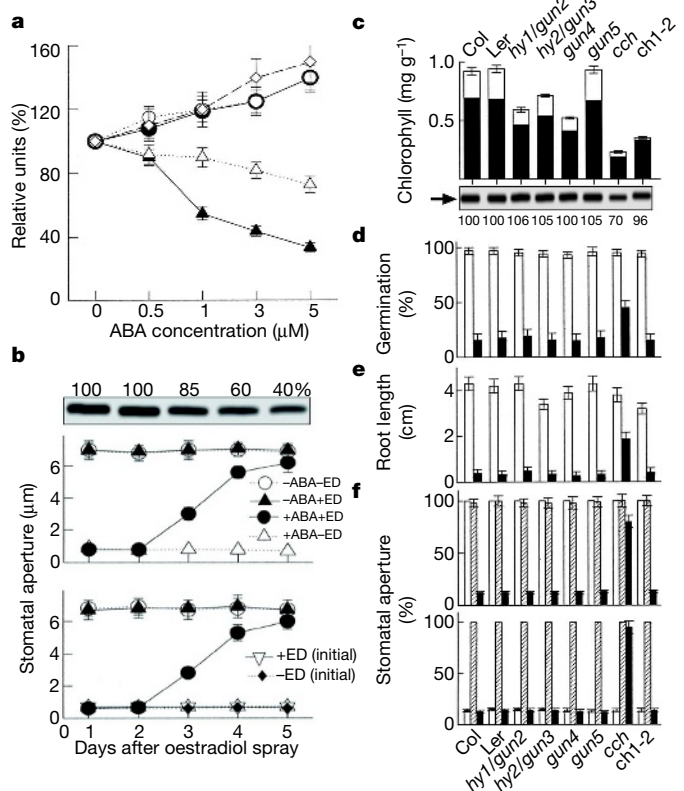


Figure 3 | ABAR-mediated ABA signalling is a distinct process. **a**, ABA treatment decreases the chlorophyll (filled triangles) and Proto (open triangles) levels, but enhances MgProto (filled circles) and ABAR protein (diamonds) levels and Mg-chelatase (open circles) activity in gl1 seedlings. **b**, Oestradiol (ED)-induced downregulation of ABAR expression (top panel; immunosignal and its relative intensity) in the inducible RNAi plants results in ABA-insensitive phenotypes in both stomatal closure (middle panel) and inhibition of stomatal opening (bottom panel) induced by 10 μ M ABA. **c**, The ABAR level (indicated by arrow and relative band intensity) is reduced in *cch*, but not in the other mutants containing different levels of chlorophyll. Filled columns, chlorophyll *a*; open columns, chlorophyll *b*. **d–f**, *cch* is an ABA-insensitive mutant in germination (**d**, 1 μ M ABA), seedling growth (**e**, root length of 7-day-old seedlings in 10 μ M ABA) and stomatal movement (**f**, top, stomatal closure; bottom, stomatal opening; 10 μ M ABA was used). Filled columns in **d** and **e**, with ABA; open columns, without ABA. Black columns in **f**, with ABA, 2 h; grey columns, without ABA, 2 h; white columns, without ABA, 0 h. Error bars indicate s.d. ($n = 5$).

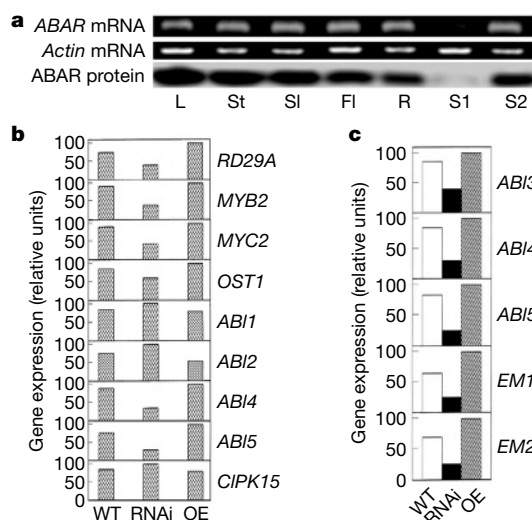


Figure 4 | Spatial expression of ABAR and alteration of ABA-signalling genes in transgenic plants. **a**, ABAR expression at the mRNA (actin mRNA as a loading control) and protein (immunodetected with anti-ABAR serum) levels in leaves (L), stems (St), siliques (SI), flowers (FI), roots (R), dry seeds (S1), and seeds kept at 20 $^{\circ}$ C for 24 h after stratification (S2). **b**, **c**, Real-time PCR analysis for expression of ABA-signalling genes in leaves (**b**) and green siliques (**c**) of gl1 (WT), RNAi (line 12) and overexpressing (OE line 1) plants.

ABA-hypersensitive phenotypes with an elevated number of ABA-binding sites; fourth, a loss-of-function mutation in *ABAR* results in an immature embryo; and last, a *cch* mutant that downregulates both *ABAR* expression and ABA-binding activity is an ABA-insensitive mutant like the post-transcriptional gene-silencing RNAi or antisense mutants. Thus, *ABAR* is a common key component in the chlorophyll biosynthetic process of chelating Mg^{2+} to Proto, plastid retrograde signalling to the nucleus and perception of the ABA signal. However, *ABAR*-mediated ABA signalling is distinct from other pathways like *ABAR*/*GUN5*-mediated plastid-to-nucleus signaling, which is independent of chlorophyll biosynthesis⁹.

As a receptor, *ABAR* regulates a series of the components involved in the ABA signalling network (Fig. 4), but the downstream components interacting directly with *ABAR* will have to be identified in the future to explain how ABA signal perception by *ABAR* is relayed in cells. *ABAR* is a single-copy gene in the *Arabidopsis* genome, is highly conserved in plant species and even shares high sequence similarities to its homologues in bacteria. This evolutionary conservation indicates a possibly vital role for it in these organisms. Gaining a further insight into how *ABAR* works in this complex signalling network will be of great interest in understanding cell signalling in plants.

METHODS

Plant materials, generation of transgenic plants and growth conditions. *Arabidopsis thaliana* ecotype gl1 was used in the generation of transgenic plants. The constructs for creating transgenic *ABAR*-RNAi, *ABAR*-antisense and *ABAR*-overexpressing lines or chemically inducible RNAi lines are described in Supplementary Information. These constructs were introduced into the GV3101 strain of *Agrobacterium tumefaciens* and transformed into plants by floral infiltration. The homozygous T₃ seeds of the transgenic plants were used for analysis. For the inducible RNAi, ten putative transgenic lines were tested by northern and western blotting after application of 10 μ M 17 β -oestradiol (Sigma), and all ten lines showed a significant decline in *ABAR* transcript and product. The seeds of the *gun4-1*, *gun5-1* and *cch* mutants were a gift from J. Chory. The seeds of the mutants *hyl-1* (ABRC number CS67), *hy2-1* (CS68) and *chl-2* (CS3362) were obtained from the Arabidopsis Biological Resource Center. Except for the mutants *hyl-1* and *hy2-1* with the ecotype Ler as background, all mutants were isolated from the ecotype Columbia. Plants were grown in a growth chamber at 20–21 °C on Murashige–Skoog medium at about 80 μ mol photons $m^{-2} s^{-1}$, or in compost soil at about 120 μ mol photons $m^{-2} s^{-1}$ over a 16-h photoperiod.

ABA binding assays. The ABA-binding activity of *ABAR* was assayed in accordance with previously described procedures⁵, with modifications. The expression of the recombinant *ABAR* in yeast, purification of the expressed *ABAR* protein, and the detailed procedures of ABA-binding assays including *in vitro* binding and pull-down assays are described in Supplementary Information.

Other assays. Screening of T-DNA insertion knockout mutants in the *ABAR* gene, phenotype analysis, RNA gel blotting, reverse transcriptase-mediated PCR, real-time PCR, production of anti-*ABAR* serum, immunoblotting and immunolabelling, chlorophyll and porphyrin measurements, assays of the effects of ABA on *ABAR* expression and Mg-chelatase activity, and induction of RNAi are described in Supplementary Information.

Received 20 May 2006; accepted 16 August 2006.

- Leung, J. & Giraudat, J. Absciscic acid signal transduction. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 199–222 (1998).
- Finkelstein, R. R., Gampala, S. S. & Rock, C. D. Absciscic acid signaling in seeds and seedlings. *Plant Cell* **14**, S15–S45 (2002).
- Himmelbach, A., Yang, Y. & Grill, E. Relay and control of absciscic acid signaling. *Curr. Opin. Plant Biol.* **6**, 470–479 (2003).
- Hornberg, C. & Weiler, E. W. High affinity binding sites for absciscic acid at the plasmalemma of *Vicia faba* guard cells. *Nature* **310**, 321–324 (1984).
- Zhang, D. P., Wu, Z. Y., Li, X. Y. & Zhao, Z. X. Purification and identification of a 42-kilodalton absciscic acid-specific-binding protein from epidermis of broad bean leaves. *Plant Physiol.* **128**, 714–725 (2002).
- Razem, F. A., Luo, M., Liu, J. H., Abrams, S. R. & Hill, R. D. Purification and characterization of a barley aleurone absciscic acid-binding protein. *J. Biol. Chem.* **279**, 9922–9929 (2004).
- Razem, F. A., El-Kereamy, A., Abrams, S. R. & Hill, R. D. The RNA-binding protein FCA is an absciscic acid receptor. *Nature* **439**, 290–294 (2006).
- Walker, C. J. & Willows, R. D. Mechanism and regulation of Mg-chelatase. *Biochem. J.* **327**, 321–333 (1997).

- Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, N. & Chory, J. *Arabidopsis* genomes uncoupled 5 (*GUN5*) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl Acad. Sci. USA* **98**, 2053–2058 (2001).
- Surpin, M., Larkin, R. M. & Chory, J. Signal transduction between the chloroplasts and the nucleus. *Plant Cell* **14**, S327–S338 (2002).
- Strand, Å., Asami, T., Alonso, J., Ecker, J. R. & Chory, J. Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature* **421**, 79–83 (2003).
- Nott, A., Jung, H. S., Koussevitzky, S. & Chory, J. Plastid-to-nucleus retrograde signaling. *Annu. Rev. Plant Biol.* **57**, 730–759 (2006).
- Giraudat, J. et al. Isolation of the *Arabidopsis* *ABI3* gene by positional cloning. *Plant Cell* **4**, 1251–1261 (1992).
- Nambara, E., Naito, S. & McCourt, P. A mutant of *Arabidopsis* which is defective in seed development and storage protein accumulation is a new *abi3* allele. *Plant J.* **2**, 435–441 (1992).
- Karger, G. A., Reid, J. D. & Hunter, C. N. Characterization of the binding of deuteroporphyrin IX to the magnesium chelatase H subunit and spectroscopic properties of the complex. *Biochemistry* **40**, 9291–9299 (2001).
- Chamovitz, D., Pecker, I. & Hirschberg, J. The molecular basis of resistance to the herbicide norflurazon. *Plant Mol. Biol.* **16**, 967–974 (1991).
- Guo, H. S., Fei, J. F., Xie, Q. & Chua, N. H. A chemical-regulated inducible RNAi system in plants. *Plant J.* **34**, 383–392 (2003).
- Yamaguchi-Shinozaki, K. & Shinozaki, K. A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell* **6**, 251–264 (1994).
- Abe, H. et al. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcription activators in absciscic acid signaling. *Plant Cell* **15**, 63–78 (2003).
- Leung, J. et al. *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. *Science* **264**, 1448–1452 (1994).
- Meyer, K., Leube, M. P. & Grill, E. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264**, 1452–1455 (1994).
- Gosti, F. et al. *ABI1* protein phosphatase 2C is a negative regulator of absciscic acid signaling. *Plant Cell* **11**, 1897–1909 (1999).
- Leung, J., Merlot, S. & Giraudat, J. The *Arabidopsis* *ABSCISIC ACID-INSENSITIVE2* (*ABI2*) and *ABI1* encode homologous protein phosphatase 2C involved in absciscic acid signal transduction. *Plant Cell* **9**, 759–771 (1997).
- Davies, S., Kurepa, J. & Vierstra, R. D. The *Arabidopsis thaliana* *HY1* locus, required for phytochrome-chromophore biosynthesis, encodes a protein related to heme oxygenases. *Proc. Natl Acad. Sci. USA* **96**, 6541–6546 (1999).
- Muramoto, T., Kohchi, T., Yokota, A., Hwang, I. & Goodman, H. M. The *Arabidopsis* photomorphogenic mutant *hyl* is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase. *Plant Cell* **11**, 335–347 (1999).
- Kohchi, T. et al. The *Arabidopsis* *HY2* gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. *Plant Cell* **13**, 425–436 (2001).
- Larkin, R. M., Alonso, J. M., Ecker, J. R. & Chory, J. *GUN4*, a regulator of chlorophyll synthesis and intracellular signalling. *Science* **299**, 902–906 (2003).
- Verdecia, M. A. et al. Structure of the Mg-chelatase cofactor *GUN4* reveals a novel hand-shaped fold for porphyrin binding. *PLoS Biol.* **3**, 777–789 (2005).
- Espineda, C. E., Linford, A. S., Devine, D. & Brusslan, J. A. The *AtCAO* gene, encoding chlorophyll *a* oxygenase, is required for chlorophyll *b* synthesis in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 10507–10511 (1999).
- Finkelstein, R. R. Mutations at two new *Arabidopsis* ABA response loci are similar to the *abi3* mutations. *Plant J.* **5**, 765–771 (1994).
- Finkelstein, R. R., Wang, M. L., Lynch, T. J., Rao, S. & Goodman, H. M. The *Arabidopsis* absciscic acid response locus *ABI4* encodes an APETALA2 domain protein. *Plant Cell* **10**, 1043–1054 (1998).
- Finkelstein, R. R. & Lynch, T. J. The *Arabidopsis* absciscic acid response gene *ABI5* encodes a basic leucine zipper transcription factor. *Plant Cell* **12**, 599–609 (2000).
- Mustilli, A. C., Merlot, S., Vavasseur, A., Fenzi, F. & Giraudat, J. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by absciscic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14**, 3089–3099 (2002).
- Guo, Y. et al. A calcium sensor and its interacting protein kinase are global regulators of absciscic acid signaling in *Arabidopsis*. *Dev. Cell* **3**, 233–244 (2002).
- Manfre, A. J., Lanni, L. M. & Marcotte, W. R. Jr. The *Arabidopsis* group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiol.* **140**, 140–149 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. H. Chua for the CLX inducible RNAi system, and Z. Z. Gong and X. C. Wang for advice and help on materials. This work was funded by the National Key Basic Research '973' Program of China and National Natural Science Foundation of China (to D.P.Z.).

Author Information The sequence of the cDNA encoding part of CHLH is deposited in GenBank under accession number DQ376081. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.P.Z. (zhangdp@sohu.net).

Spatially regulated ubiquitin ligation by an ER/nuclear membrane ligase

Min Deng¹ & Mark Hochstrasser¹

The ubiquitin system targets many cellular proteins. Doa10 (also known as Ssm4), a yeast transmembrane ubiquitin ligase (E3), resides in the endoplasmic reticulum (ER), but it attaches ubiquitin to soluble proteins that concentrate in the nucleus. A central question is how nuclear substrates gain access to an enzyme in the ER. Here we show that Doa10 reaches the inner nuclear membrane. A subcomplex of nuclear pore subunits is important for this transport. Notably, another ER transmembrane E3, Hrd1 (also known as Der3), cannot localize efficiently to the inner nuclear membrane. Tethering Doa10 at the cell periphery inhibits degradation of soluble nuclear substrates but not cytoplasmic ones. If Doa10 is released from these peripheral sites, localization of Doa10 to the nuclear envelope and degradation of its nuclear substrates are restored in parallel. Thus, localization of Doa10 to the inner nuclear membrane is necessary for nuclear substrate degradation. These data indicate that different membrane ubiquitin ligases are spatially sorted within the ER–nuclear envelope membrane system and that this differential localization contributes to their specificity.

Protein degradation by the ubiquitin–proteasome pathway is responsible for most cellular protein turnover^{1–3}. Among the principal mediators of substrate specificity are the ubiquitin ligases (E3s), which, in conjunction with one or more specific ubiquitin-conjugating enzymes (E2s), recognize protein determinants called degradation signals or degrons⁴. Substrates are modified by ubiquitin or polymeric ubiquitin chains; polyubiquitin-modified proteins can then be recognized and degraded by the 26S proteasome³. Mammalian cells have hundreds of different E3s, whereas the model eukaryote *Saccharomyces cerevisiae* is estimated to have 60–100. E3s and E2s often localize to specific cellular compartments, and such compartmentalization can be a key component of their specificity⁵.

Doa10 is a conserved, 151-kDa integral membrane E3 with 14 transmembrane segments^{6,7}. Together with two ER-associated E2s, Ubc6 and Ubc7, Doa10 modifies a broad range of substrates, including ER membrane proteins, soluble nuclear proteins and soluble cytoplasmic substrates^{6,8,9}. One nuclear substrate, the Mat α 2 transcription factor, is targeted to the Doa10 pathway by its Deg1 degron^{8,10}. How Mat α 2 and other nuclear substrates gain access to the ER-localized Doa10 has remained a mystery. Either nuclear substrates are exported out of the nucleus to reach Doa10, or the E3 is transported to the inner nuclear membrane (INM) to reach its nuclear substrates (see Supplementary Fig. S1). As our preliminary data argued against the first possibility, we focused on investigating the latter hypothesis.

Targeted silencing indicates that Doa10 is in the INM

Immunogold electron microscopy was attempted for localizing Doa10–GFP (green fluorescent protein; Supplementary Fig. 2). Doa10–GFP, which is fully functional⁶, showed apparent localization to both sides of the nuclear envelope. However, endogenous Doa10 levels are very low, and only a few thin sections had bead distributions that appeared to show labelling on both the inner and outer nuclear envelope.

Given the limitations of these data, we developed two assays as alternatives to ultrastructural localization. The first was a ‘targeted silencing’ assay¹¹, which was initially designed to study the role of

nuclear localization in gene silencing. Here we adapted it to determine whether Doa10 expressed at endogenous levels could localize to the INM. The yeast silent mating-type locus gene silencer *HMR-E* contains three DNA elements (A, E and B), which recruit silencing factors (Fig. 1a). Deletion of any two elements leads to full de-repression of the *HMR* locus, but silencing is restored by anchoring the locus to the nuclear periphery where silencing factors are concentrated^{11,12}. By fusing the DNA-binding domain of Gal4 (G_{BD}) to Doa10, we measured potential INM tethering of an *HMR* derivative that had two silencing elements replaced by a triplicated Gal4-binding site (UAS_G). Expression of G_{BD} –Doa10 and endogenous Doa10 was similar (a 13×Myc epitope tag was appended to both) (Fig. 1b).

With an *hmr::TRP1* reporter, which allows gene expression to be measured by growth on medium lacking tryptophan, silencing was seen in the tester strain expressing G_{BD} –Doa10 (Fig. 1c, Supplementary Fig. 3). The effect was slightly less pronounced than with the positive control, overexpressed G_{BD} –Stt3 (ref. 11), but depended on the presence of the UAS_G sites and required at least one natural silencer sub-element as well as the Sir2 silencing factor (Supplementary Fig. 3). Therefore, G_{BD} –Doa10-mediated repression exhibited all the hallmarks of targeted silencing. We also detected specific binding of G_{BD} –Doa10 to the UAS_G sites in the modified *hmr::TRP1* locus by chromatin immunoprecipitation (Fig. 1d, lanes 1 and 2). These data indicate that Doa10 can reach the inner nuclear envelope.

Differential localization of ER membrane E3s

For a broader analysis of integral membrane protein localization to the INM, we developed a second, quantifiable cell-based assay that exploits the observation that increased expression of the Nup53 nucleoporin causes specific proliferation of the INM¹³. The resulting intranuclear membrane lamellae pack against the nuclear envelope (see Supplementary Fig. 2, right) and often also cut across the nuclear interior¹³. If Doa10 is in the INM, visualizing Doa10 in cells that overproduce Nup53 should reveal similar structures.

The *DOA10-GFP* strain was transformed with a *CUP1-NUP53* plasmid. After induction with copper, intranuclear lamellae were observed by electron microscopy. Both confocal (Fig. 2a) and

¹Yale University, Department of Molecular Biophysics and Biochemistry, 266 Whitney Avenue, P.O. Box 208114, New Haven, Connecticut 06520-8114, USA.

conventional fluorescence microscopy (Fig. 2b) showed that Nup53 overexpression led to the appearance of a distinctive Doa10–GFP signal that transected the nucleus in ~25% of the cells (Fig. 2f); the nuclei resembled the Greek letter theta (θ), so we refer to them as ‘theta nuclei’. A similar frequency of theta nuclei was observed when Nup53 was tagged with GFP (Fig. 2f). Neither the control vector nor *CUP1–NUP53 Δ C*, which fails to drive INM proliferation¹³, induced the appearance of theta nuclei (Fig. 2; Supplementary Fig. 4). We did not see theta nuclei in cells that overexpressed Hmg1 (HMG-CoA reductase), which induces ER lamellae around the outside of the nucleus (‘karmellae’)¹⁴ (Supplementary Fig. 4).

To validate the Nup53-based assay, we examined two proteins as controls: Sec61, an integral ER membrane protein that is found in intranuclear membranes in certain mammalian cells¹⁵, and Mga2, a membrane-anchored transcription factor precursor that is predicted to be excluded from the INM^{16,17}. Theta nuclei were seen with Sec61–GFP in ~25% of Nup53-overexpressing cells (Fig. 2c, f), but virtually no such nuclei were observed with Mga2–GFP (Fig. 2d, f). Therefore, the fluorescence assay provides a simple means for testing localization of membrane proteins to the INM, including those that also occupy the ER and outer envelope.

In contrast to Doa10, the other main transmembrane E3 of the yeast ER, Hrd1 (ref. 18), showed little if any localization to the INM (Fig. 2e, f), indicating that INM protein localization is highly selective. Three other proteins that function with Doa10—Ubc6, Ubc7 and Cue1—revealed frequencies of theta nuclei similar to Doa10–GFP (Fig. 2f). Doa10 is thought to form a membrane-embedded complex

with these proteins, and the localization data support the possibility that such a complex exists in the INM. Our results show that the Doa10 and Hrd1 E3 complexes concentrate in different subdomains within the continuous ER–nuclear envelope membrane system.

Role of the nuclear pore complex

Little is known about the transport of polytopic membrane proteins to the inner nuclear envelope. INM proteins are postulated to move from the ER through the lipid bilayer to the INM through lateral channels in the nuclear pore complex (NPC)^{19–21}. Initial evidence that the NPC might be involved in this process came from micro-injection of animal cells with antibodies or lectins that bind NPC subunits; this had modest and variable effects on INM localization of a model membrane protein²². To test the involvement of the NPC in INM transport and to gain insight into which components of the NPC are important for this, we examined Doa10–GFP by the Nup53-based assay in a panel of nucleoporin mutants (Fig. 2g). Importantly, deletion of two specific nucleoporins, Pom152 and Nup188, partially but significantly reduced the formation of theta nuclei. This was not due to inhibition of the import of Nup53, as Nup53–GFP showed normal frequencies of theta nuclei in *pom152 Δ* and *nup188 Δ* mutants (Fig. 2f). As expected, theta nuclei were almost never observed in *nup170 Δ* cells (Fig. 2g) because Nup170 is required for nuclear import of Nup53 (ref. 13).

Pom152 is an integral membrane protein, and it therefore could be in close proximity to, or a component of, the NPC lateral channels. Moreover, Pom152 directly binds Nup188, and a *pom152 Δ nup188 Δ* double mutation is lethal^{20,23,24}. The most straightforward interpretation of our data is that Nup188 and Pom152 form part of an NPC structure that is needed for efficient membrane protein movement through the NPC. If INM localization of Doa10 is necessary for efficient targeting of its nuclear substrates, the reduction in INM-localized Doa10, albeit modest, might lead to detectable defects in nuclear protein degradation. Indeed, degradation of *Deg1*– β gal, a nuclear substrate of Doa10, was moderately but significantly

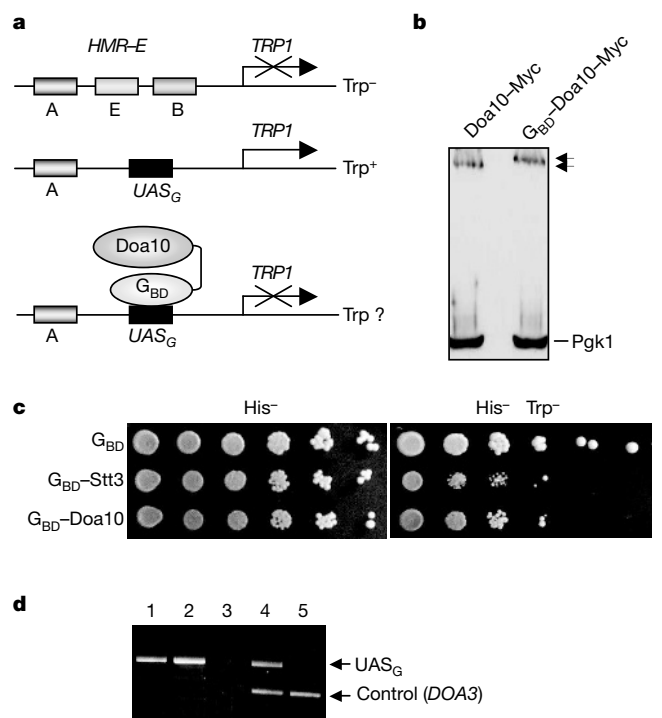


Figure 1 | Localization of a transmembrane ubiquitin ligase to the inner nuclear envelope. **a**, Targeted-silencing assay. In wild-type cells, *hmr::TRP1* is repressed (Trp^-). With a triplicated *UAS_G* replacing the E and B sites of the silencer (*Aeb::UAS_G::TRP1*; strain YSB35), silencing is lost (Trp^+). Introducing a *G_{BD}*–membrane protein fusion restores silencing if the protein can anchor the locus to the inner nuclear envelope¹¹. **b**, *G_{BD}*–Doa10–Myc expression is similar to chromosomal Doa10–Myc expression. Pgk1 is a 45-kDa globular protein. **c**, Targeted silencing by *G_{BD}*–Doa10 assayed with serially diluted YSB35 cells on drop-out plates lacking histidine (for plasmid selection) or both histidine and tryptophan. *G_{BD}*–Stt3 is the positive control. **d**, Chromatin immunoprecipitation analysis of *G_{BD}*–Doa10 binding to *UAS_G* in YSB35 (lanes 1–4) and YSB1 (no *UAS_G*, lane 5) cells. Anti-*G_{BD}* (lane 1), anti-Myc (lane 2) or no antibody (lane 3) was used. Lanes 4, 5: input extracts of YSB35 and YSB1 cells, respectively.

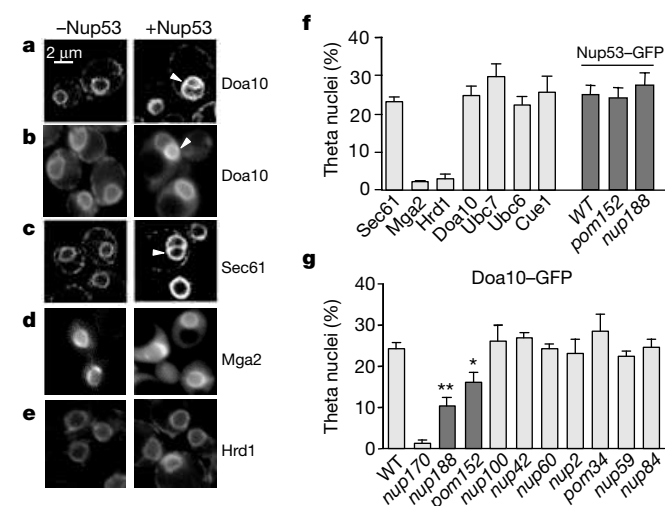


Figure 2 | Fluorescence assay of INM protein localization. **a–e**, *pCUP1–NUP53* (right column) or vector (left column) was transformed into cells with chromosomally expressed GFP fusions to Doa10 (**a**, **b**), Sec61 (**c**) or Hrd1 (**e**), or into *ufd1-1* cells harboring a *GFP–Mga2* plasmid (**d**); the *ufd1-1* mutation prevents cleavage of latent Mga2 but does not affect INM trafficking. Cells were examined by confocal (**a**, **c**) or standard fluorescence microscopy (**b**, **d**, **e**). Arrowheads mark theta nuclei. **f**, Percentage of theta nuclei in cells expressing the indicated protein–GFP fusions. Error bars in **f**, **g**: standard deviation of three independent cultures. **g**, Doa10–GFP transport to the INM measured in nucleoporin deletion strains as percentage of theta nuclei. Double asterisk, $P < 0.001$ (relative to wild type, WT); single asterisk, $P < 0.01$.

impaired in *pom152Δ* and *nup188Δ* strains (Supplementary Fig. 5), although we cannot rule out indirect effects of the mutations.

Function of INM localization of Doa10

To test the inference that localization of Doa10 to the INM is functionally relevant, we tethered Doa10 to the cell periphery by fusing it to the actin-binding domain of coronin (Crn1) (Supplementary Fig. 6). Crn1 binds actin filaments and localizes to cortical actin patches²⁵. The first 400 residues of Crn1 were fused to the carboxy terminus of Doa10, followed by GFP; this *doa10-CNG* allele replaced the chromosomal *DOA10* gene. Doa10-CNG was expressed as a full-length protein (data not shown) and localized to foci that were primarily at the cell cortex (Fig. 3a), resembling the localization of native Crn1. Most importantly, Doa10-CNG failed to localize to the nuclear

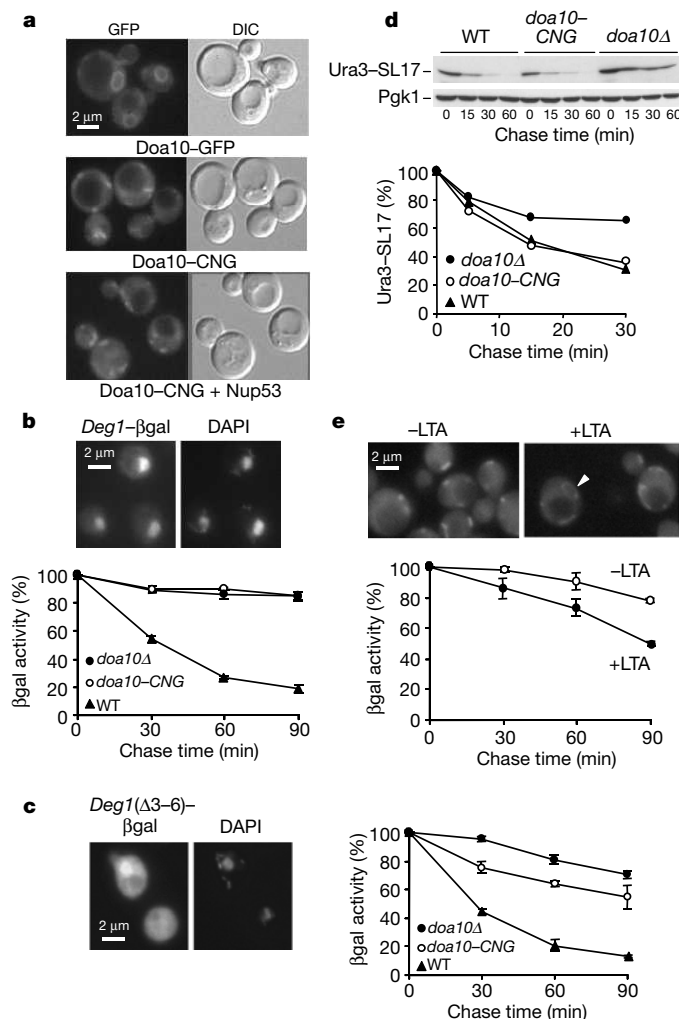


Figure 3 | Localization of Doa10 to the INM is required for nuclear substrate degradation. **a**, Doa10-CNG localizes to foci at the cell periphery; this is unaltered by Nup53 overexpression. **b**, Cytoplasmically tethered Doa10-CNG is inactive in nuclear *Deg1-βgal* degradation. Error bars in **b**, **c**, **e**: standard deviations of three independent measurements. Nuclear localization of *Deg1-βgal* was confirmed by anti-βgal immunofluorescence (top). **c**, The *Deg1(Δ3-6)-βgal* mutant is partially impaired in nuclear localization, which correlates with enhanced degradation in *doa10-CNG* cells relative to *doa10Δ* ($P < 0.05$). **d**, A soluble cytoplasmic substrate is degraded normally in *doa10-CNG* cells by cycloheximide-chase/immunoblot (upper panel) and standard pulse-chase (lower panel) analyses of the cytoplasmic HA-tagged Ura3-SL17 substrate. **e**, LTA treatment partially restores localization to the nuclear envelope and nuclear substrate degradation. *doa10-CNG* cells were treated with DMSO or 100 μM LTA. Arrowhead indicates nuclear envelope localization. Degradation of *Deg1-βgal* was significantly enhanced in LTA-treated cells ($P < 0.01$).

envelope and did not form theta nuclei when Nup53 was overexpressed. Proteolysis of the nuclear Doa10 substrate *Deg1-βgal* was severely impaired in *doa10-CNG* cells (Fig. 3b). The *doa10-CNG* allele did not cause a general ER stress response or defects in ER-associated degradation as measured by a *UPRE-lacZ* reporter assay and degradation of the misfolded ER substrate carboxypeptidase Y (CPY*) (data not shown).

When the nuclear localization signal (NLS) in *Deg1* was mutated²⁶, a substantial fraction of the resulting *Deg1(Δ3-6)-βgal* protein localized to the cytoplasm (Fig. 3c). A corresponding increase in degradation of *Deg1(Δ3-6)-βgal* was observed in *doa10-CNG* cells compared with *doa10Δ* cells. These results support the idea that cytoplasmically tethered Doa10 is at least partially functional but cannot access nuclear-localized substrate. Consistent with this, degradation of a Doa10 substrate that is almost exclusively cytoplasmic, Ura3-SL17 (ref. 9), occurred at wild-type rates when measured either by cycloheximide-chase/immunoblot (Fig. 3d, top) or pulse-chase analysis (bottom). Therefore, Doa10-CNG is functional against soluble cytosolic substrates but not nuclear ones.

When *doa10-CNG* cells were treated with latrunculin-A (LTA), an actin filament assembly inhibitor²⁵, actin patches disassembled within 5 min (not shown). Relocalization of Doa10-CNG from cortical sites to the nuclear envelope was much slower and incomplete, either because Doa10-CNG localization was stabilized by additional cortical interactions or because the Doa10-Crn1 fusion tended to aggregate²⁷. Nevertheless, we detected a weak nuclear rim signal with Doa10-CNG in ~20% of cells after 2.5 h in LTA (Fig. 3e). Importantly, in parallel with this partial restoration of localization to the nuclear envelope, we observed a statistically significant recovery of *Deg1-βgal* degradation (Fig. 3e).

Ubiquitin modification of the naturally short-lived Matα2 repressor occurs through two pathways, one that depends on Ubc6-Ubc7-Doa10, and the other on Ubc4 (refs 6, 8). Matα2 proteolysis was inhibited around threefold in both *doa10-CNG* and *doa10Δ* cells, indicating that nucleus-excluded Doa10-CNG is not active against Matα2 (Fig. 4a). Similarly, in both *doa10Δ ubc4Δ* and *doa10-CNG ubc4Δ* double mutants, Matα2 had the same half-life (~53 min), confirming that Doa10-CNG was not active against MATα2. Notably, partial restoration by LTA of Doa10-CNG localization at the nuclear envelope was associated with partial but significant suppression of the Matα2 degradation defect ($t_{1/2} \approx 30$ min; Fig. 4b). This was true even though LTA slightly inhibited Matα2 turnover in wild-type cells. We conclude that localization of Doa10 to the inner nuclear envelope is required for efficient targeting of a natural nuclear substrate.

Determinants of Doa10 INM localization

To address whether Doa10 contains specific INM-targeting sequences, we made complementary deletions, which removed either C-terminal residues 951-1319 (Doa10-N) or N-terminal residues 2-950 (Doa10-C), and performed theta-nuclei assays on the resulting GFP fusion proteins. Both proteins localized normally to the INM, indicating that Doa10 does not bear one unique INM-targeting or retention signal (Fig. 4c). However, it might carry two or more redundant signals.

Doa10 might be able to diffuse into the INM because it lacks any cytoplasmic loops larger than ~24 kDa. By contrast, the Hrd1 C-terminal cytoplasmic domain is ~38 kDa in size, and this might not allow passage through the NPC lateral channels²¹. However, fusion of the Hrd1 cytoplasmic domain to Doa10 did not block INM accumulation (Fig. 4c), indicating that the exclusion of Hrd1 from the INM depends on its membrane domain. Attachment of a larger globular protein, Pgk1 (~45 kDa), to Doa10 did significantly attenuate INM localization (Fig. 4c), consistent with a size limit that restricts diffusion through the NPC. The ability of these Doa10 fusion proteins to access the INM correlated with their ability to support degradation of Doa10 nuclear substrates (Supplementary Fig. 7a),

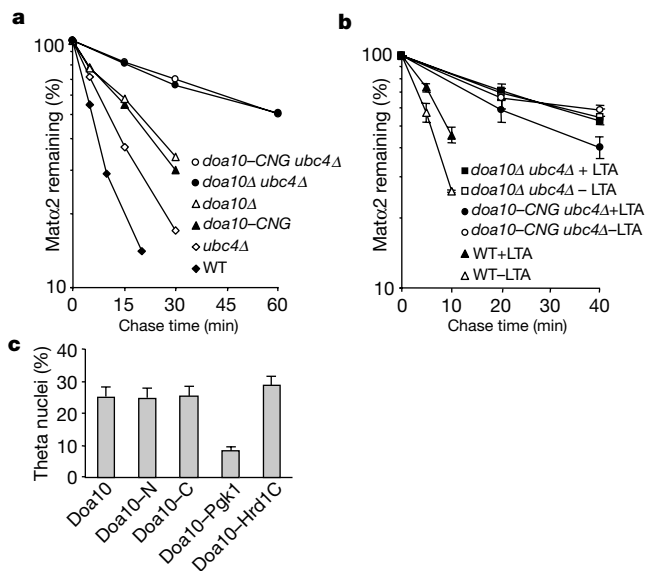


Figure 4 | Nuclear substrate degradation and determinants of Doa10 INM localization. **a**, Pulse-chase analysis of Mat α 2 degradation. Calculated half-lives of Mat α 2 were 6, 13, 19, 19, 52 and 53 min in wild-type, *ubc4 Δ* , *doa10 Δ* , *doa10-CNG*, *doa10-CNG ubc4 Δ* and *doa10 Δ ubc4 Δ* cells, respectively. **b**, Partial recovery of Mat α 2 degradation through LTA treatment. Error bars in **b**, **c**: standard deviation of three independent cultures. LTA treatment of *doa10-CNG ubc4 Δ* cells significantly increased Mat α 2 degradation rates ($P < 0.05$) even though LTA impaired degradation in wild-type (WT) cells 1.7-fold ($P < 0.01$). **c**, Effects of Doa10 deletions and protein-domain fusions on Doa10 movement to the INM.

whereas degradation of cytosolic substrates was minimally affected by either fusion protein (Supplementary Fig. 7b).

The spatial regulation of ubiquitin system enzymes is increasingly recognized as a key aspect of their substrate specificity⁵. We have shown here that the selective localization of a transmembrane E3 complex to the INM is required for degradation of its nuclear substrates. This is the first example of a polytopic ER–nuclear envelope enzyme whose enzyme activity has been shown to be required in the nucleus. Spatial control of ubiquitin conjugation in the nucleus might be intricate. There is evidence for controlled movement of specific proteins and chromosomal loci to the nuclear periphery in both yeast and higher cells^{28–30}. Localization of a ubiquitin ligase to the inner nuclear envelope would allow it to transfer ubiquitin to chromosomal and nuclear proteins specifically when they move to the nuclear periphery. This provides a potential mechanism for the temporal and spatial regulation of transcription and other nuclear processes by ubiquitin modification.

Note added in proof: A very recent study³⁴ reported a pathway for INM trafficking of two related yeast transmembrane proteins that concentrate in the INM. This pathway is NLS-dependent and requires the Nup2 nucleoporin but not Nup188 or Pom152. In contrast, Doa10 has no obvious NLS, and its trafficking to the INM depends on Nup188 and Pom152 but not Nup2. These findings indicate that there are multiple pathways for integral membrane protein trafficking to the INM. For a protein such as Doa10, which is also needed in the ER, a pathway that strongly concentrates it in the INM should not be favoured.

METHODS

Yeast media and methods. Cells were grown at 30 °C in either rich (YPD) or selective minimal media. Yeast *hmr::TRP1* silencing assays were done by spotting tenfold serial dilutions of overnight cultures onto selective plates. The *G_{BD}* fusion alleles were carried on plasmids that also had a *HIS3* marker³¹. Growth was assessed after two days at 30 °C. For assays of theta-nuclei formation, induction of Nup53 was achieved by growing cells carrying the *CUP1-NUP53* plasmid in selective medium in the presence of 100 μ M copper sulphate for 16 h (ref. 13).

Degradation assays. Cycloheximide-chase/immunoblot analysis to follow Ura3–SL17 (haemagglutinin (HA)-epitope tagged) degradation was performed by growing strains to optical density at 600 nm (OD_{600}) ~ 1 in selective medium containing 100 μ M CuSO₄ followed by addition of cycloheximide to 0.5 mg ml^{–1}. Culture aliquots (2 OD_{600} equivalents) were removed at the desired time points. After incubating the aliquots in 100 μ M NaOH at room temperature for 5 min, cells were pelleted and resuspended in Laemmli SDS gel loading buffer and heated at 100 °C for 10 min. Immunoblotting was done with an anti-HA monoclonal antibody, 16B12 (Covance), and anti-Pgk1 (Molecular Probes) was added to provide a loading control. Pulse-chase experiments to measure the half-lives of Ura3–SL17 and Mat α 2 were performed as described previously⁸. Half-lives were calculated after phosphorimager quantitation. Quantitative activity-chase analyses of Deg1– β gal degradation were done by blocking protein synthesis with cycloheximide and measuring loss of β gal activity with ortho-nitrophenyl galactoside (ONPG) substrate³². For statistical comparison of degradation rates, half-lives were calculated by linear regression, and mean values were compared by unpaired *t*-tests (GraphPad InStat).

Immunoelectron microscopy. Electron microscopy (EM) was performed by the University of Colorado EM facility as described³³. Affinity-purified rabbit polyclonal antibody to GFP diluted 1:100 in blocking solution was followed by goat anti-rabbit secondary antibody conjugated to 15 nm colloidal gold (1:20 dilution in blocking buffer).

Received 30 June; accepted 15 August 2006.

- Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439 (1996).
- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Pickart, C. M. & Eddins, M. J. Ubiquitin: structures, functions, mechanisms. *Biochim. Biophys. Acta* **1695**, 55–72 (2004).
- Varshavsky, A. Regulated protein degradation. *Trends Biochem. Sci.* **30**, 283–286 (2005).
- Pines, J. & Lindon, C. Proteolysis: anytime, any place, anywhere? *Nature Cell Biol.* **7**, 731–735 (2005).
- Swanson, R., Locher, M. & Hochstrasser, M. A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Mat α 2 repressor degradation. *Genes Dev.* **15**, 2660–2674 (2001).
- Kreft, S. G., Wang, L. & Hochstrasser, M. Membrane topology of the yeast endoplasmic reticulum-localized ubiquitin ligase Doa10 and comparison with its human ortholog TEB4 (MARCH-VI). *J. Biol. Chem.* **281**, 4646–4653 (2006).
- Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MAT α 2 repressor. *Cell* **74**, 357–369 (1993).
- Ravid, T., Kreft, S. G. & Hochstrasser, M. Membrane and soluble substrates of the Doa10 ubiquitin ligase are degraded by distinct pathways. *EMBO J.* **25**, 533–543 (2006).
- Johnson, P. R., Swanson, R., Rakhilina, L. & Hochstrasser, M. Degradation signal masking by heterodimerization of MAT α 2 and MAT α 1 blocks their mutual destruction by the ubiquitin-proteasome pathway. *Cell* **94**, 217–227 (1998).
- Andrulis, E. D., Neiman, A. M., Zappulla, D. C. & Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394**, 592–595 (1998); erratum **395**, 525 (1998).
- Taddei, A., Hediger, F., Neumann, F. R. & Gasser, S. M. The function of nuclear architecture: a genetic approach. *Annu. Rev. Genet.* **38**, 305–345 (2004).
- Marelli, M., Lusk, C. P., Chan, H., Aitchison, J. D. & Wozniak, R. W. A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *J. Cell Biol.* **153**, 709–724 (2001).
- Wright, R., Basson, M., D'Ari, L. & Rine, J. Increased amounts of HMG-CoA reductase induce “karmellae”: a proliferation of stacked membrane pairs surrounding the yeast nucleus. *J. Cell Biol.* **107**, 101–114 (1988).
- Isaac, C., Pollard, J. W. & Meier, U. T. Intranuclear endoplasmic reticulum induced by Nopp140 mimics the nucleolar channel system of human endometrium. *J. Cell Sci.* **114**, 4253–4264 (2001).
- Hitchcock, A. L. *et al.* The conserved Npl4 protein complex mediates proteasome-dependent membrane-bound transcription factor activation. *Mol. Biol. Cell* **12**, 3226–3241 (2001).
- Rape, M. *et al.* Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48^{UFD1/NPL4}, a ubiquitin-selective chaperone. *Cell* **107**, 667–677 (2001).
- Hampton, R. Y. ER-associated degradation in protein quality control and cellular regulation. *Curr. Opin. Cell Biol.* **14**, 476–482 (2002).
- Hinshaw, J. E., Carragher, B. O. & Milligan, R. A. Architecture and design of the nuclear pore complex. *Cell* **69**, 1133–1141 (1992).
- Suntharalingam, M. & Wente, S. R. Peering through the pore: nuclear pore complex structure, assembly, and function. *Dev. Cell* **4**, 775–789 (2003).
- Worman, H. J. & Courvalin, J. C. The inner nuclear membrane. *J. Membr. Biol.* **177**, 1–11 (2000).

22. Ohba, T., Schirmer, E. C., Nishimoto, T. & Gerace, L. Energy- and temperature-dependent transport of integral proteins to the inner nuclear membrane via the nuclear pore. *J. Cell Biol.* **167**, 1051–1062 (2004).
23. Nehrbass, U., Rout, M. P., Maguire, S., Blobel, G. & Wozniak, R. W. The yeast nucleoporin Nup188p interacts genetically and physically with the core structures of the nuclear pore complex. *J. Cell Biol.* **133**, 1153–1162 (1996).
24. Damelin, M. & Silver, P. A. In situ analysis of spatial relationships between proteins of the nuclear pore complex. *Biophys. J.* **83**, 3626–3636 (2002).
25. Goode, B. L. et al. Coronin promotes the rapid assembly and cross-linking of actin filaments and may link the actin and microtubule cytoskeletons in yeast. *J. Cell Biol.* **144**, 83–98 (1999).
26. Hall, M. N., Hereford, L. & Herskowitz, I. Targeting of *E. coli* β -galactosidase to the nucleus in yeast. *Cell* **36**, 1057–1065 (1984).
27. Appleton, B. A., Wu, P. & Wiesmann, C. The crystal structure of murine coronin-1: a regulator of actin cytoskeletal dynamics in lymphocytes. *Structure* **14**, 87–96 (2006).
28. Ishii, K., Arib, G., Lin, C., Van Houwe, G. & Laemmli, U. K. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* **109**, 551–562 (2002).
29. Casolari, J. M. et al. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* **117**, 427–439 (2004).
30. Brickner, J. H. & Walter, P. Gene recruitment of the activated *INO1* locus to the nuclear membrane. *PLoS Biol.* **2**, e342 (2004).
31. Ma, J. & Ptashne, M. A new class of yeast transcriptional activators. *Cell* **51**, 113–119 (1987).
32. Plemper, R. K., Bohmler, S., Bordallo, J., Sommer, T. & Wolf, D. H. Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. *Nature* **388**, 891–895 (1997).
33. Giddings, T. H. Jr et al. Using rapid freeze and freeze-substitution for the preparation of yeast cells for electron microscopy and three-dimensional analysis. *Methods Cell Biol.* **67**, 27–42 (2001).
34. King, M. C., Lusk, C. P. & Blobel, G. Karyopherin-mediated import of integral inner nuclear membrane proteins. *Nature* **442**, 1003–1007 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Sternglanz and R. Wozniak for strains and plasmids, P. Crews for generously providing LTA, O. Kerscher for Supplementary Fig. 1, and J. Wu and J. Wolenski for advice on confocal microscopy. We are grateful to S. Kreft, A. Lewis and T. Ravid for comments on the manuscript. This work was supported by the NIH.

Author Contributions M.D. performed all the experiments, and M.D. and M.H. planned the experiments, performed the data analysis and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (mark.hochstrasser@yale.edu).

LETTERS

An almost head-on collision as the origin of two off-centre rings in the Andromeda galaxy

D. L. Block¹, F. Bournaud^{2†}, F. Combes², R. Groess¹, P. Barmby³, M. L. N. Ashby³, G. G. Fazio³, M. A. Pahre³ & S. P. Willner³

The unusual morphology of the Andromeda galaxy (Messier 31, the closest spiral galaxy to the Milky Way) has long been an enigma. Although regarded for decades as showing little evidence of a violent history, M31 has a well-known^{1–7} outer ring of star formation at a radius of ten kiloparsecs whose centre is offset from the galaxy nucleus. In addition, the outer galaxy disk is warped, as seen at both optical⁸ and radio⁹ wavelengths. The halo contains numerous loops and ripples. Here we report the presence of a second, inner dust ring with projected dimensions of 1.5×1 kiloparsecs and offset by about half a kiloparsec from the centre of the galaxy (based upon an analysis of previously-obtained data¹⁰). The two rings appear to be density waves propagating in the disk. Numerical simulations indicate that both rings result from a companion galaxy plunging through the centre of the disk of M31. The most likely interloper is M32. Head-on collisions between galaxies are rare, but it appears nonetheless that one took place 210 million years ago in our Local Group of galaxies.

Newly acquired images¹⁰ of M31 secured by the Infrared Array Camera¹¹ (IRAC) onboard the Spitzer Space Telescope span the wavelength regime of 3.6–8.0 μm . These images offer unique probes of the morphologies of the stellar distribution and interstellar medium with no interference from extinction. Figure 1 shows the emission map of the interstellar medium at 8 μm , generated by

subtracting a scaled 3.6 μm image (dominated by starlight) from the 8 μm image. The subtraction removes the contribution from stellar photospheres and leaves only the emission from dust grains¹⁰, which trace the interstellar medium of M31.

What is most striking in Fig. 1 (and in the enlarged inset) is the presence of a complete—although asymmetric—inner ring of dust 6.9 by 4.4 arcmin in extent, translating to linear dimensions of about 1.5 by 1 kpc (assuming a distance¹² of 780 kpc). The inner ring lies between the two well-known Baade spiral dust arms¹³, both of which are clearly seen in emission. The inner ring is elongated in a direction close to the minor axis and belongs to the central gas disk, which appears to be more face-on¹⁴. It is therefore not possible to know the inner ring's precise ellipticity, but it is unlikely to be circular.

IRAC imaging thus reveals two rings. The outer ring is offset by approximately 10% of its radius, while the inner ring is offset by about 40% or ~ 0.5 kpc. The inner elliptical ring has been alluded to in earlier studies^{15,16}, but all investigators have hitherto believed it to be a mini-spiral, related to a bar. Published Spitzer 24 μm images⁵ of M31 show centrally concentrated dust emission; the ring morphology is therefore disguised at these longer wavelengths. The IRAC images beautifully show the inner ring at high spatial resolution and furthermore confirm that this feature is a complete and continuous ring, even though it is offset and asymmetrical.

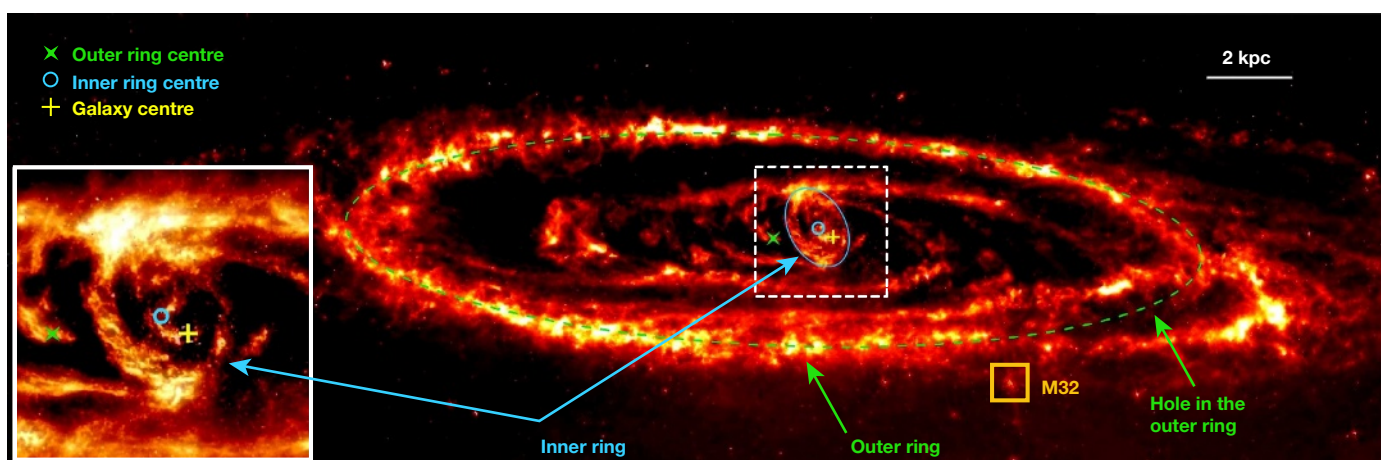


Figure 1 | The Andromeda galaxy M31 observed with IRAC¹¹ on board the Spitzer Space Telescope at a wavelength of 8 μm . A scaled version of the 3.6 μm image was subtracted from the 8 μm image to remove light from stellar photospheres. The remaining emission therefore traces the emission of warm dust grains and macromolecules in the interstellar medium. M31 clearly possesses two rings. Apart from the famous outer dust ring seen at a

radius of ~ 10 kpc, this map reveals a second 1.5 kpc by 1 kpc inner dust ring offset by approximately 0.5 kpc from the galaxy nucleus. Both rings are interpreted to be density waves induced by an almost head-on collision. The most likely candidate is the dwarf companion galaxy M32, which appears faint in this image because it has little dust. The starlight of M32 is, however, prominent in the separate 3.6 and 8 μm images¹⁰.

¹Anglo American Cosmic Dust Laboratory, School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa.

²Observatoire de Paris, LERMA, 61 Avenue de l'Observatoire, F-75014 Paris, France. ³Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [†]Present address: DSM/DAPNIA/Service d'Astrophysique, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France.

There are two known scenarios whereby disk systems form rings: by head-on galaxy collisions or by rotating bars. Head-on collisions differ from common tidal interactions between galaxies because the pericentric distance of the orbit of the companion is small, and its orbit is almost perpendicular to the target disk. Such collisions produce expanding ring-shaped waves¹⁷. Rotating bars in spiral disks produce inner, outer and sometimes nuclear rings in barred spiral galaxies. The rings occur at orbital resonances and arise from the galaxy's internal dynamics (unlike collisional rings). Many examples of bar-induced rings are known¹⁸.

Infrared imaging of spiral galaxies in the local Universe confirms the ubiquity of the bar phenomenon; about 80% of spiral galaxies are barred¹⁹, even if many are classified as unbarred at visible wavelengths. M31 is no exception; although Hubble classified it as normal type Sb, it is likely to possess a stellar bar. The presence of a bar in M31 is suggested by its boxy bulge, which is conspicuous in infrared imaging^{5,20} and is aligned almost parallel to the galaxy's major axis. The boxy bulge might be an inflated bar, or it may hide a thinner parallel bar (at about 10° from the major axis). There is, however, no bar parallel to the minor axis, along which the major axis of the inner dust ring lies. Had there been a stellar bar along the minor axis, it would surely have been detected in the IRAC 3.6 μm image¹⁰, which is much deeper than the 2.2 μm Two-Micron All-Sky Survey (2MASS) images and less affected by dust attenuation. The double-ring system of M31 therefore appears to be unrelated to any possible central bar, because the elliptical inner ring does not have the typical orientation of most bar-induced gas rings (which are generally aligned along the bar's major axis). Moreover, the off-centre ring strongly supports the collision interpretation. Resonant rings generated from rotating bars are rarely as off-centre as the rings of M31, and the amount of deceleration increases with radius. In M31, however, the inner ring is off-centre by 40% of its radius, whereas the outer ring is off-centre by only 10%.

The relative brightness of the inner and outer rings further supports a collision origin. When a bar induces a strong pair of rings, the most prominent one is generally the inner, located just beyond the extremity of the bar at the so-called 4:1 resonance. In M31, the outer ring is much brighter than the inner one. This makes the interpretation of bar-induced resonance rings highly unlikely except under the unrealistic hypothesis of a very slowly rotating bar extending up to the outer ring (contrary to observations). The dual-ring morphology of M31 strongly indicates expanding density waves triggered by a head-on galaxy collision with a companion.

The dwarf companion galaxy M32 is very likely to be the impactor. To investigate whether a collision of M32 with the disk of M31 could induce the two ring-shaped density waves identified in the IRAC images, *N*-body simulations were performed²¹. The simulations include stars, gas and dark matter in M31 and M32 with one million particles and a spatial resolution of 350 parsecs. Our simulations differ from all previous ones in that the initial orbit of M32 is close to the polar axis of M31, and the mass ratio is larger. Previous simulations⁵ have assumed that the impact occurred several kiloparsecs away from the centre of M31 and not along the rotation axis (nearly head-on). Our new simulations start with an axisymmetric disk, proceed for approximately one gigayear while a bar and spiral arms form, then introduce a collision with M32. We take into account not only the mass stripped during the collision but also the dark matter associated with M32 itself. We assume an initial mass ratio for M32 of a tenth that of M31 (including dark matter). Current-epoch mass estimates of M32 are lower, but much mass is likely to have been stripped²² during the collision. The final mass of M32 in our model is 1/23 that of M31, which is compatible with present-day mass estimates.

The simulation of a head-on collision (Fig. 2) beautifully reproduces the observed global morphologies of M31 reported in this Letter. The model produces two striking density wave rings offset from the galaxy centre. The outer density wave propagates through

the M31 disk, resulting in a perturbed interlacing of spiral features with an outer ring. This produces the conspicuous hole observed in the outer 10 kpc ring of M31 (compare Figs 1 and 2). Our model reproduces this very striking (but transient) feature and adds evidence that the two rings have been induced by a collision 210 million years ago close to the rotation axis of M31. The non-zero (but small) impact parameter makes the inner ring off-centre, and the inclination of the orbit with respect to the rotation axis produces the ring elongation. In the simulations, the head-on collision does not destroy the pre-existing boxy bulge in the barred target disk but merely weakens it.

Collisional ring galaxies are expected to show local radial and tangential perturbed motions in the vicinity of the rings, which will be observed as 'streaming motions'. When the intruder is a quarter the mass of the disk galaxy or less, the ring behaves like a wave which propagates through the disk²³, as opposed to the large-scale bulk

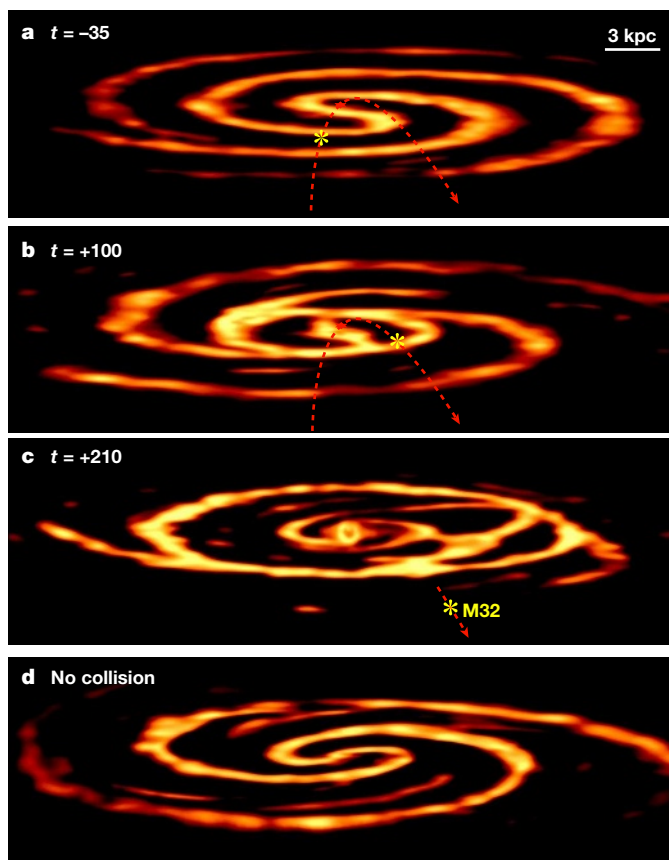


Figure 2 | Gas morphology produced by simulations²¹ of a head-on encounter between M31 and M32. These *N*-body models include the gravitational dynamics of stars, interstellar matter, and dark matter in both M31 and M32. A 'sticky-particle' scheme accounts for the dissipative nature of the interstellar medium, and star formation is also included²¹. The dashed red line demarcates the orbit of M32; the locale of impact lies very close to the polar axis of M31. Snapshots **a**, **b** and **c** occur at $t = 35$ million years before collision and at 100 and 210 million years post-impact; **c** also shows the position of M32 as we see it today. All snapshots are taken with the disk viewed at an inclination angle of 77° for direct comparison with Fig. 1. **c** shows the central region of M31 warped by a tilt angle of 30° with respect to the main disk plane in accord with observations¹⁴. The two ring-like waves (both offset from the galaxy centre) are seen, as is the hole in the outer ring. **d** shows the gas morphology at $t = 210$ million years, starting from the same initial disk conditions but modelled without a collision; no density-wave rings are generated. The initial mass of the companion M32 is a tenth that of M31 (or 1/13th excluding dark matter). The companion is assumed to have struck the M31 disk at a velocity of 265 km s^{-1} and 1.2 kpc from the centre of M31. It would now be located at a distance of 35 kpc and at a galactic latitude of about 45° , which is fully compatible with the present position³⁰ of M32.

motion of material that occurs when the mass of the intruder is larger. Our model predicts radial velocities in the gas of 10 km s^{-1} at the outer (10 kpc) ring. Unwin²⁴ found local streaming velocities in neutral hydrogen gas of 30 km s^{-1} at that radius; our predicted radial velocities constitute 30% of the total observed streaming velocity. The inner ring is predicted to show lesser radial velocities, which are more difficult to probe because of the vertical tilt of the central gas disk. There is an interesting kinematic signature of a ring induced by collision: the resulting positive radial velocity in the ring is predicted to be associated with a depression in the rotational velocity towards the outer parts of the ring (with a corresponding increase in the inner parts of the ring). This is indeed seen in isovelocity contours in neutral hydrogen²⁴. The predicted tangential streaming motions for spiral arm density waves²⁵ would have the opposite sense. (See the Supplementary Information.)

The most striking analogue of the disk morphology observed in M31 is the Cartwheel galaxy. The Cartwheel galaxy is the archetype of a double-ringed morphology produced by collision. Two distinct density wave rings are observed: a conspicuous outer ring, associated with massive star formation, and an elliptical inner ring, offset from the centre of the galaxy. In between the two rings lie several spiral arms, termed 'spokes', which have developed in a trailing pattern. Several models have been computed of the Cartwheel galaxy, simulating the almost head-on collision with one of its companions^{26,27}.

M31 differs from the Cartwheel galaxy in that we do not see the inner ring in starlight but only in gas and dust. In other words, the inner ring in the Cartwheel galaxy shows a much greater degree of contrast (compared to the parent disk) than does the inner ring in M31. The explanation is that the impactor in the Cartwheel galaxy must have been more massive than M32, resulting in two density-wave rings of stars. Our simulations match the lower contrast seen in M31: the primary ring wave propagates outward in the disk, being created by a crowding of particle trajectories and made prominent by massive star formation triggered at the peak of the wave. The second elongated density-wave ring begins to propagate in the central regions of the galaxy, most probably in a tilted central disk. At the present epoch, the gas morphology of the model agrees with the double-ring structure of the M31 interstellar medium as revealed by IRAC.

The morphology of M31 has been mysterious for many years, but the discovery of the offset inner ring may be the clue needed to offer an explanation: a recent head-on collision could have produced both the inner ring and the previously known outer one. The rings are unlikely to have been created by bar resonances because of the rings' relative brightness and their offsets from the galaxy centre and also because they show no relationship to the spiral structure; in particular, no minor axis bar is observed. While head-on collisions between galaxies may have been common in the early Universe^{28,29}, only a handful are known nearby. The discovery of one in our near-neighbour M31 affords the unique opportunity of studying such a collision at unprecedented spatial resolution.

Received 24 May; accepted 18 August 2006.

1. Brinks, E. & Shane, W. W. A high resolution hydrogen line survey of Messier 31. I. Observations and data reduction. *Astron. Astrophys. Suppl.* **55**, 179–251 (1984).
2. Dame, T. M., Koper, E., Israel, F. P. & Thaddeus, P. A complete CO survey of M31. I. Distribution and kinematics. *Astrophys. J.* **418**, 730–742 (1993).
3. Nietten, Ch, et al. Molecular gas in the Andromeda galaxy. *Astron. Astrophys.* **453**, 459–475 (2006).
4. Haas, M. et al. Cold dust in the Andromeda Galaxy mapped by ISO. *Astron. Astrophys.* **338**, L33–L36 (1998).
5. Gordon, K. D. et al. Spitzer MIPS Infrared Imaging of M31: Further evidence for a spiral-ring composite structure. *Astrophys. J. Lett.* **638**, L87–L92 (2006).

6. Beck, R., Berkhuijsen, E. M. & Hoernes, P. A deep $\lambda 20 \text{ cm}$ radio continuum survey of M 31. *Astron. Astrophys.* 329–336 (1998).
7. Pellet, A. et al. A survey of H II regions in M31. *Astron. Astrophys.* 439–461 (1978).
8. Ibata, R. et al. A giant stream of metal-rich stars in the halo of the galaxy M31. *Nature* **412**, 49–51 (2001).
9. Braun, R. The distribution and kinematics of neutral gas in M31. *Astrophys. J.* **372**, 54–66 (1991).
10. Barmby, P. et al. Dusty waves on a starry sea: the mid-infrared view of M31. *Astrophys. J. Lett.* (in the press); preprint at (<http://arXiv.org/astro-ph/0608593>) (2006).
11. Fazio, G. G. et al. The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J. Suppl.* **154**, 10–17 (2004).
12. Stanek, K. Z. & Garnavich, P. M. Distance to M31 with the Hubble Space Telescope and HIPPARCOS red clump stars. *Astrophys. J. Lett.* **503**, L131–L134 (1998).
13. Baade, W. *Evolution of Stars and Galaxies* 56–68 (Harvard Univ. Press, Cambridge, 1963).
14. Jacoby, G. H., Ford, H. & Ciardullo, R. Ionized gas in the center of M31. *Astrophys. J.* **290**, 136–139 (1985).
15. Willaime, M.-C., Lequeux, J., Melchior, A.-L. & Hanus, M. in *The Promise of the Herschel Space Observatory* (eds Pilbratt, G. L. et al.) 519–521 (ESA-SP-460, European Space Agency, Noordwijk, 2001).
16. Ciardullo, R. et al. The morphology of the ionized gas in M31's bulge. *Astron. J.* **95**, 438–444 (1988).
17. Appleton, P. N. & Struck-Marcell, C. Collisional ring galaxies. *Fundamentals Cosmic Phys.* **16**, 111–220 (1996).
18. Buta, R. The catalog of southern ringed galaxies. *Astrophys. J. Suppl.* **96**, 39–116 (1995).
19. Block, D. L. et al. The gravitational torque of bars in optically unbarred and barred galaxies. *Astron. Astrophys.* **375**, 761–769 (2001).
20. Beaton, R. L. et al. Unveiling the boxy bulge and bar of the Andromeda Spiral Galaxy. *Astrophys. J. Lett.* (submitted).
21. Bournaud, F. & Combes, F. Formation of polar ring galaxies. *Astron. Astrophys.* **401**, 817–833 (2003).
22. Welch, G. A. & Sage, L. J. The Interstellar Medium of M32. *Astrophys. J.* **557**, 671–680 (2001).
23. Gerber, R. A., Lamb, S. A. & Balsara, D. S. A stellar and gas dynamical numerical model of ring galaxies. *Mon. Not. R. Astron. Soc.* **278**, 345–366 (1996).
24. Unwin, S. C. Neutral hydrogen in the Andromeda Nebula—III. The velocity field. *Mon. Not. R. Astron. Soc.* **205**, 773–786 (1983).
25. Shu, F. H., Milione, V. & Roberts, W. W. Nonlinear gaseous density waves and galactic shocks. *Astrophys. J.* **183**, 819–842 (1973).
26. Struck-Marcell, C. & Higdon, J. L. Hydrodynamic models of the Cartwheel ring galaxy. *Astrophys. J.* **411**, 108–124 (1993).
27. Horellou, C. & Combes, F. A model for the Cartwheel ring galaxy. *Astrophys. Space Sci.* **276**, 1141–1149 (2001).
28. Lavery, R. J., Remijan, A., Charmandaris, V., Hayes, R. D. & Ring, A. A. Probing the evolution of the galaxy interaction/merger rate using collisional ring galaxies. *Astrophys. J.* **612**, 679–689 (2004).
29. Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S. & Schaffer, M. A. Galaxy morphologies in the Hubble Ultra Deep Field: dominance of linear structures at the detection limit. *Astrophys. J.* **631**, 85–100 (2005).
30. Mateo, M. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Funding for this work was provided by the Anglo American Chairman's Fund as well as by NASA through an award issued by JPL/Caltech.

Author Contributions P.B., G.G.F., M.L.N.A., M.A.P. and S.P.W. provided the IRAC data of M31 on which the analyses are based. D.L.B. initiated the collaboration between the USA, France and WITS, South Africa and identified Baade dust spirals in emission in the central region of M31. F.C. discovered the central ring between the inner Baade arms and believed it to be induced by a head-on collision. F.B. performed the numerical simulations which fully corroborated this interpretation and also generated Figs 1 and 2. R.G. cleaned the IRAC images of foreground stars and conducted Fourier spectral analysis of the Baade arms. F.C. and F.B. provided crucial drafts of this Letter. The final revised versions were prepared by D.L.B., S.P.W., R.G. and M.L.N.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L.B. (block@cam.wits.ac.za).

No evidence for thick deposits of ice at the lunar south pole

Donald B. Campbell¹, Bruce A. Campbell², Lynn M. Carter², Jean-Luc Margot¹ & Nicholas J. S. Stacy³

Shackleton crater at the Moon's south pole has been suggested as a possible site of concentrated deposits of water ice, on the basis of modelling of bi-static radar polarization properties and interpretations of earlier Earth-based radar images^{1,2}. This suggestion, and parallel assumptions about other topographic cold traps, is a significant element in planning for future lunar landings. Hydrogen enhancements have been identified in the polar regions³, but these data do not identify the host species or its local distribution. The earlier Earth-based radar data lack the resolution and coverage for detailed studies of the relationship between radar scattering properties, cold traps in permanently shadowed areas, and local terrain features such as the walls and ejecta of small craters. Here we present new 20-m resolution, 13-cm-wavelength radar images that show no evidence for concentrated deposits of water ice in Shackleton crater or elsewhere at the south pole. The polarization properties normally associated with reflections from icy surfaces in the Solar System^{4–6} were found at all the observed latitudes and are strongly correlated with the rock-strewn walls and ejecta of young craters, including the inner wall of Shackleton. There is no correlation between the polarization properties and the degree of solar illumination. If the hydrogen enhancement observed by the Lunar Prospector orbiter³ indicates the presence of water ice, then our data are consistent with the ice being present only as disseminated grains in the lunar regolith.

The possible presence of water-ice deposits in the polar areas of the Moon has been a controversial issue since the mid-1990s. The 1.6° inclination of the Moon's rotation axis to the normal to the ecliptic plane means that there are areas near the poles, primarily the floors and lower interior walls of impact craters, that are in permanent shadow from the Sun. Several researchers^{7,8} have pointed out that ice could be stable at the low temperatures (<100 K) expected in these shadowed areas, and the idea was given significant impetus by the discovery of probable ice deposits at the poles of Mercury by Earth-based radars^{9,10}. The radar echoes from Mercury's poles were interpreted as reflections from water ice on the basis of their similarity to the unusual properties of radar echoes from the icy Galilean satellites of Jupiter⁴. Low-temperature water-ice surfaces can exhibit a very strong opposition effect; that is, they preferentially scatter the incident energy back towards the radar. They also preferentially reflect the same sense of circular polarization that was transmitted, leading to a circular polarization ratio (CPR, the ratio of the reflected power in the same circular (SC) sense of polarization transmitted to that in the opposite (OC) sense) greater than unity. A mirror-like reflection would have a CPR of zero, and most geological surfaces have a CPR of less than unity¹¹. It is thought that these properties of the radar reflection are related to the very small propagation loss of low-temperature water ice by means of a coherent backscatter opposition effect (CBOE), the coherent addition of a ray that

propagates into the ice, is scattered and emerges in the direction of the radar, with its path-reversed twin¹². Laboratory experiments using lasers on assemblages of particles indicate that volume scattering in a low-loss medium is not necessarily required; the CBOE can also be exhibited by a rough scattering surface¹³. The polar deposits on Mercury are thought to be water ice or some other condensate¹⁴ because they are located only on those parts of impact crater floors that are in permanent shadow¹⁵.

The first reports that there might be deposits of water ice at the lunar south pole arose from bi-static radar observations of very low spatial resolution in 1994, with the 13-cm-wavelength telemetry system of the Clementine orbiter as a transmitter and one of the 70-m NASA/Deep Space Network antennas to receive the echo¹. The south pole was observed at a viewing angle of 4.5–5.5° (the angle of the 70-m DSN antenna above the horizon at the lunar south pole) over a small range of bi-static phase angles centred on zero (that is, on opposition). A ~25% increase in the CPR was observed as the bi-static angle passed through zero. This increase was observed only for the pass over the south pole and was interpreted as indicating the probable presence of low-loss volume scatterers such as water ice¹. Subsequent analysis of the bi-static data led to the suggestion that ~10 km² of ice might be present on the lower, permanently shadowed, part of the Earth-facing inner slope of the 19-km-diameter south pole crater, Shackleton². This location was selected to be coincident with an area of high CPR discovered in 125-m-resolution radar imagery obtained in 1992 with the 13-cm radar system on the US National Science Foundation's (NSF's) Arecibo telescope in Puerto Rico¹⁶. The Arecibo observations were made at a south pole viewing angle of 6.1°. However, in addition to the Earth-facing inner wall of Shackleton crater, many small, probably fresh, craters exhibited CPRs with values ranging from 1.0 to 2.4 (ref. 16). Similar CPR values were observed for several small areas in Sinus Iridum that are clearly sunlit (at latitude ~47° N), leading to the suggestion that the high CPRs were due to scattering in very rough or blocky terrain¹⁶.

To investigate further the radar scattering properties of the terrain at the south pole of the Moon, and of Shackleton crater in particular, we have made radar imaging observations at spatial resolutions as fine as 20 m. In April and October 2005 we used the Arecibo telescope to transmit a right circularly polarized signal at 13 cm wavelength and received the echo in both senses of circular polarization with the NSF's 100-m Robert C. Byrd Green Bank Telescope (GBT) in West Virginia, USA. The GBT is ~2,300 km from Arecibo, corresponding to a bi-static angle of 0.37°. Results of our observations on 16 April and 24 October 2005, when the viewing angle for the south pole was close to maximum at ~6.5°, are reported here. The April data provided imagery, with a high signal-to-noise ratio, at 20 m resolution, significantly higher resolution than for any previous observations. However, the relative calibration between the channels was poor,

¹Department of Astronomy, Cornell University, Ithaca, New York 14853, USA. ²Center for Earth and Planetary Studies, Smithsonian Institution, Box 37012, Washington, DC 20013-7012, USA. ³Defence Science and Technology Organization, Box 1500, Edinburgh, SA 5111, Australia.

making absolute measurements of the CPR difficult. Well-calibrated 100-m-resolution data from October were used to calibrate the CPR values obtained from the April data.

Figure 1a is the OC radar backscatter image from October covering the south pole and terrain on the lunar nearside north to about 68° S. The CPR values overlaid on this image are shown in Fig. 1b. The geology of the south-pole region is dominated by highland terrain formed by overlapping ejecta from nearby craters and the major basins. Shackleton (Fig. 2a) seems to be younger than neighbouring craters such as de Gerlache, Shoemaker or Faustini, and probably dates to the Eratosthenian period, consistent with earlier geological mapping¹⁷. Low-lying areas between the highland massifs (for example the region to the west of de Gerlache), and many of the radar-visible older crater floors, are characterized by smooth-textured terrain that has been mapped as basin-related ejecta, most probably from Orientale^{18,19}. These deposits have high CPRs relative to the surrounding terrain at both 13-cm (Fig. 1b) and 70-cm wavelengths¹⁸, indicating a greater abundance of centimetre-scale to decimetre-scale rocks within the upper few metres of the regolith.

Figure 1b shows many locations with CPRs ranging from 1.0 to 1.5 or more. The measured values of the CPR are within 20% of those obtained from the 1992 lower-spatial-resolution 13-cm imaging¹⁶. As was postulated on the basis of the lower-resolution data¹⁶, there is a clear correlation between high values of the CPR and the walls and ejecta deposits of both large and small craters. These high values raise two questions: first, are they in any way indicative of the presence of water ice, and second, if not water ice, then what combination of surface morphology and scattering mechanism is responsible?

Shackleton crater (Fig. 2a) is of considerable interest for NASA's lunar exploration programme because of the suggested presence of water ice on its Earth-facing inner slope², and because a few locations

on its rim may be permanently sunlit for periods of up to 200 days each year, making them suitable for extended human habitation²⁰. The upper portion of Shackleton's Earth-facing inner rim can be seen in both Lunar Orbiter and Clementine visible imagery, indicating that it is sunlit at least part of the time; any detected ice deposits would have to be located on areas lower down on the wall. Our new, high-resolution data (Fig. 2b) show that there is no systematic variation in the CPR with distance down the (radar) visible part of the inner slope, and high CPR values occur in seasonally sunlit regions of the crater wall and the near-rim ejecta blanket.

Shackleton is morphologically similar to the 17 km crater Schomberger G (77.1° S, 7.7° E; Fig. 2c). Both seem to have layered outcrops, or terraces, within the upper few hundred metres of their interior rim deposits, which could make access to the interior walls very challenging. The CPR values for the inner Earth-facing slope of Schomberger G (Fig. 2d), which is sunlit during the lunar diurnal cycle, are almost identical to those of Shackleton. The inner and outer walls and floor of the nearby 31-km Copernican-period crater Schomberger A (78.8° S, 24.4° E) have CPR values of 1.5 and greater (Fig. 1b), as do many other smaller young craters. It is clear that CPR values greater than 1 do not require the presence of water ice and are therefore not necessarily indicative of the presence of ice. For Schomberger A and G, and numerous other craters, the high ratios must instead be associated with scattering from the rocks and blocky material that comprise the proximal ejecta or that have cascaded down the steep inner slopes. This scattering could exhibit an opposition effect that, like its optical counterpart, includes a component arising from a CBOE. High CPR values are not confined to the Moon. They have also been observed for three near-Earth asteroids with measured values between 1.0 and 1.2 (ref. 21), for Maxwell Montes on Venus with similar values²², and for the SP lava flow near Flagstaff,

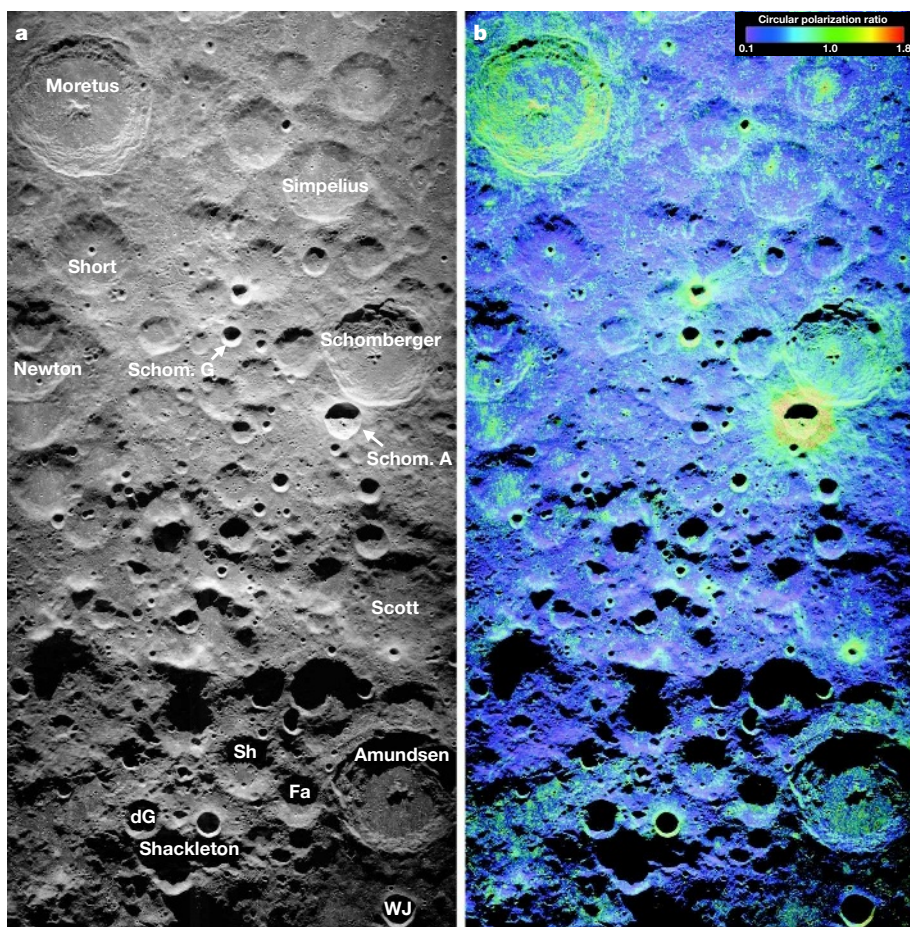


Figure 1 | Radar image data from 24 October 2005, for a region covering the south pole and the nearside to latitude $\sim 68^\circ$ S. The south pole is on the left rim of Shackleton. North is 3.5° anticlockwise from the centreline of this image (0° longitude grazes the east rim of Moretus). Spatial resolution 100 m per pixel; polar stereographic projection. **a**, OC radar image at 100 m resolution. Major craters labelled: Sh, Shoemaker; Fa, Faustini; dG, de Gerlache; WJ, Wiechert J. **b**, CPR at 500 m resolution presented as a colour overlay on the radar backscatter image. Note the very high CPRs for Schomberger A and many other smaller craters.

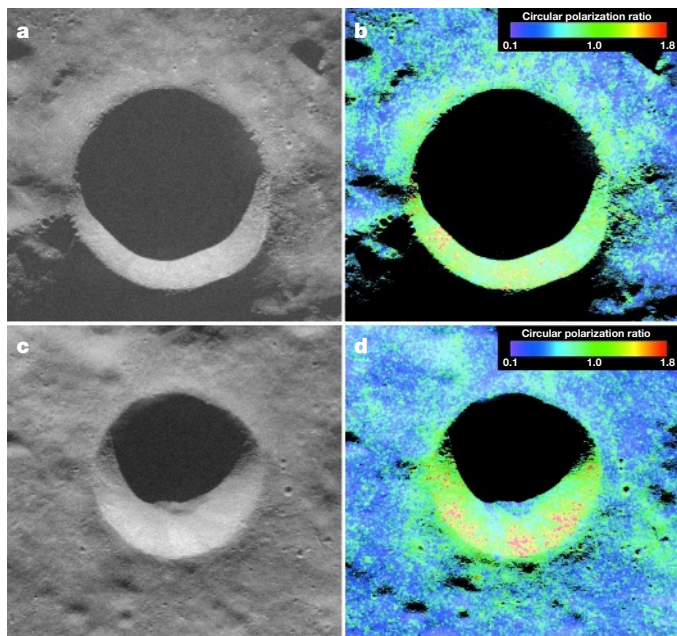


Figure 2 | Comparison of the radar-scattering properties of Shackleton and Schomberger G craters from the 16 April 2005 data. **a**, The SC radar image at 20 m resolution for the 19-km diameter Shackleton crater at an incidence angle of $\sim 83.5^\circ$. **c**, A similar image for the 17-km Schomberger G crater at an incidence angle of $\sim 70^\circ$. The two craters seem to be of similar ages. Terraced structures can be seen in the top few hundred metres of the interior deposits, and the inner wall of Schomberger G shows evidence for down-slope mass wasting. **b**, **d**, The CPRs for Shackleton (**b**) and Schomberger G (**d**) at 220 m resolution overlaid on the radar images. The CPRs have been adjusted to be identical to the calibrated but lower-resolution values of Fig. 1 (see the text).

Arizona, USA, which has CPR values of up to 2.0 at 24 cm wavelength²³. Water ice clearly has no function in radar scattering from these surfaces.

Strong evidence that the CPR values in Shackleton are due to a component of volume-scattering ice must certainly include a high degree of correlation between the occurrence of high CPR values and the transition from seasonally sunlit to permanently shadowed portions of the crater wall. Although available topographic data²⁴ do not show the precise location of this transition, it is within the area viewed by our mapping. The area of strongest CPR within Shackleton (at about the 8 o'clock position in Fig. 2b), cited as supporting the existence of a CBOE effect in the shadowed terrain on the basis of the lower-resolution 1992 data², is resolved here into a collection of high-CPR points that occur in both permanent shadow and seasonal illumination. In the absence of any strong difference in properties with solar illumination, the possibility that these features are linked with rocky debris, as observed at Schomberger G and elsewhere, remains the more likely explanation.

Neutron spectrometer measurements from the Lunar Prospector orbiter in 1998 indicated the presence of significant concentrations of hydrogen in the lunar regolith at the south pole⁵. Assuming that the hydrogen is in the form of water molecules, modelling²⁵ shows that the neutron measurements are consistent with a concentration of $1.5 \pm 0.8\%$ by mass of ice in the upper 1 m or so of the lunar regolith. If this ice is distributed as grains in the regolith at this concentration, or locally at higher concentrations, then it would not be observable with radar. If it is in the form of deposits of ice that are decimetres to 1 m or more thick, then it might be observable with radar by means of the CBOE.

Our high-resolution radar data show no evidence that high CPR values in Shackleton, or elsewhere in the south polar region, are correlated with solar illumination conditions. Rather, these high CPR values are associated with the rugged inner walls and proximal

ejecta of impact craters. This is consistent with the presence of any water ice being only as disseminated grains at 1–2% abundance²⁵. Any planning for future exploitation of hydrogen at the Moon's south pole should be constrained by this low average abundance rather than by the expectation of localized deposits at higher concentrations.

Received 3 April 2006; accepted 15 August 2006.

- Nozette, S. *et al.* The Clementine bi-static radar experiment. *Science* **274**, 1495–1498 (1996).
- Nozette, S. *et al.* Integration of lunar polar remote sensing data sets: Evidence for ice at the lunar south pole. *J. Geophys. Res.* **106**, 23253–23266 (2001).
- Feldman, W. C. *et al.* Fluxes of fast and epithermal neutrons from Lunar Prospector: Evidence for water ice at the lunar poles. *Science* **281**, 1496–1500 (1998).
- Campbell, D. B., Chandler, J. F., Ostro, S. J., Pettengill, G. H. & Shapiro, I. I. Galilean satellites—1976 radar results. *Icarus* **34**, 254–267 (1978).
- Harmon, J. K., Arvidson, R. E., Guinness, E. A., Campbell, B. A. & Slade, M. A. Mars mapping with delay-Doppler radar. *J. Geophys. Res.* **104**, 14065–14090 (1999).
- Black, G. J. & Campbell, D. B. Rhea's surface: Ice properties measured by radar. *Bull. Am. Astron. Soc.* **36**, 1123 (2004).
- Watson, K., Murray, B. C. & Brown, H. The behavior of volatiles on the lunar surface. *J. Geophys. Res.* **66**, 3033–3045 (1961).
- Arnold, J. R. Ice at the lunar poles. *J. Geophys. Res.* **84**, 5659–5668 (1979).
- Slade, M. A., Butler, B. J. & Muhleman, D. O. Mercury radar imaging: Evidence for polar ice. *Science* **258**, 635–640 (1992).
- Harmon, J. K. & Slade, M. A. Radar mapping of Mercury: Full disk images and polar anomalies. *Science* **258**, 640–643 (1992).
- Campbell, B. A. *Radar Remote Sensing of Planetary Surfaces* (Cambridge Univ. Press, Cambridge, 2002).
- Hapke, B. Coherent backscatter and the radar characteristics of the outer planet satellites. *Icarus* **88**, 407–417 (1990).
- Nelson, R. M., Hapke, B. W., Smythe, W. D. & Horn, L. J. Phase curves of selected particulate materials: The contribution of coherent backscattering to the opposition surge. *Icarus* **131**, 223–230 (1998).
- Sprague, A. L., Hunten, D. M. & Ladders, K. Sulfur at Mercury, elemental at the poles and sulfides in the regolith. *Icarus* **118**, 211–215 (1995).
- Harmon, J. K., Perillat, P. J. & Slade, M. A. High-resolution radar imaging of Mercury's north pole. *Icarus* **149**, 1–15 (2001).
- Stacy, N. J. S., Campbell, D. B. & Ford, P. G. Arecibo radar mapping of the lunar poles: A search for ice deposits. *Science* **276**, 1527–1530 (1997).
- Wilhelms, D. E., Howard, K. A. & Wilshire, H. G. *USGS Map I-1162* (US Geological Survey, Washington DC, 1979).
- Campbell, B. A. & Campbell, D. B. Regolith properties in the south polar region of the Moon from 70-cm radar polarimetry. *Icarus* **180**, 1–7 (2006).
- Campbell, B. A. & Hawke, B. R. Radar mapping of lunar cryptomaria east of Orientale basin. *J. Geophys. Res.* **110** (E9), E09002 (2005).
- Zuber, M. T. & Garrick-Bethell, I. What do we need to know to land on the Moon again? Perspective article. *Science* **310**, 983–985 (2005).
- Benner, L. A. M. *et al.* Radar detection of near-Earth asteroids 2062 Aten, 2101 Adonis, 3103 Eger, 4544 Xanthus, and 1992 QN. *Icarus* **130**, 296–312 (1997).
- Campbell, B. A., Campbell, D. B. & DeVries, C. Surface processes in the Venus highlands: Results from analysis of Magellan and Arecibo data. *J. Geophys. Res.* **104**, 1897–1916 (1999).
- Campbell, B. A., Arvidson, R. E. & Shepard, M. K. Radar polarization properties of volcanic and playa surfaces: Applications to terrestrial remote sensing and Venus data interpretation. *J. Geophys. Res.* **98**, 17099–17114 (1993).
- Margot, J. L., Campbell, D. B., Jurgens, R. F. & Slade, M. A. Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* **284**, 1658–1660 (1999).
- Feldman, W. C. *et al.* Evidence for water ice near the lunar poles. *J. Geophys. Res.* **106** (E10), 23231–23252 (2001).

Acknowledgements We thank the staff of the Arecibo and Green Bank Observatories for their assistance with the observations, and J. Chandler for providing the lunar ephemeris observing files. This work was supported in part by grants from the NASA Planetary Astronomy and Planetary Geology and Geophysics Programs. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the US National Science Foundation. The Green Bank Telescope is part of the National Radio Astronomy Observatory, a facility of the NSF operated under a cooperative agreement by Associated Universities, Inc.

Author Contributions The work was initiated by D.B.C., who also wrote the initial draft of the paper. D.B.C., B.A.C., L.M.C. and J.-L.M. planned and made the observations. Software development and data reduction were done by B.A.C., L.M.C. and J.-L.M. All authors participated in the interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B.C. (campbell@astro.cornell.edu).

LETTERS

Experimental purification of two-atom entanglement

R. Reichle¹†, D. Leibfried¹, E. Knill¹, J. Britton¹, R. B. Blakestad¹, J. D. Jost¹, C. Langer¹, R. Ozeri¹, S. Seidelin¹ & D. J. Wineland¹

Entanglement is a necessary resource for quantum applications—entanglement established between quantum systems at different locations enables private communication¹ and quantum teleportation², and facilitates quantum information processing³. Distributed entanglement is established by preparing an entangled pair of quantum particles in one location, and transporting one member of the pair to another location. However, decoherence during transport reduces the quality (fidelity) of the entanglement. A protocol to achieve entanglement ‘purification’ has been proposed⁴ to improve the fidelity after transport. This protocol uses separate quantum operations at each location and classical communication to distil high-fidelity entangled pairs from lower-fidelity pairs. Proof-of-principle experiments distilling entangled photon pairs have been carried out^{5–9}. However, these experiments obtained distilled pairs with a low probability of success and required destruction of the entangled pairs, rendering them unavailable for further processing. Here we report efficient and non-destructive entanglement purification⁴ with atomic quantum bits. Two noisy entangled pairs were created and distilled into one higher-fidelity pair available for further use. Success probabilities were above 35 per cent. The many applications of entanglement purification make it one of the most important techniques in quantum information processing.

Recent efforts to realize practical quantum information processing devices based on various physical systems have led to impressive new developments¹⁰. Much-anticipated applications of quantum information processing devices are quantum communication and quantum computing. Large-scale implementations require distributing each quantum bit (qubit) of an entangled pair of qubits to separate locations. In quantum communication, entangled pairs are the fundamental resource for quantum teleportation², entanglement-based quantum cryptography¹ and other protocols. They also underlie several promising schemes for quantum computing^{3,11}, are needed for fast coupling of distant qubits^{12,13} and play an important part in recent methods for achieving fault-tolerance¹⁴. In all applications, it is important to ensure high fidelity of the entangled pairs once they are distributed to their destinations. This is particularly difficult in quantum communication, because most schemes require conversion between stationary (usually material) qubits suitable for storage or manipulation and ‘flying’ (usually photonic) qubits suitable for transport.

To restore fidelity lost during transport of entangled qubits, we can use entanglement purification⁴ to distil a smaller number of high-fidelity entangled qubits. The simplest instance of the entanglement purification protocol of ref. 4 distils one entangled pair of qubits from two imperfectly entangled ones. Here the qubits at one location are compared by means of local quantum operations. The results of the comparison are shared through classical communication between locations, and if they are consistent, a higher-fidelity pair of entangled qubits is obtained. Assuming no error in the comparison and

sufficiently entangled initial pairs, the process can be iterated in multiple ‘rounds’ to obtain arbitrarily high-fidelity entangled pairs. Previous experiments, using photons, included demonstrations of entanglement purification^{6,9} and entanglement concentration^{5,7,8}, which required specific input states. However, in these experiments, entangled pairs were obtained by post-selection with a low success probability of $<10^{-3}$ per try. As two pairs are needed for an experiment, purification success probabilities were less than 10^{-6} per try. In addition, successful comparisons required destruction of entanglement, so the purified entangled pairs were not available for further use.

Here we report experiments that faithfully implemented the purification protocol proposed in ref. 4. An important feature of this protocol is that it works for all input states with sufficient fidelity with respect to the desired ideal entangled state. We tested the protocol on a family of input states that are approximately pure. Although other methods such as entanglement concentration can achieve better fidelity for our input states, this is at the cost of lower success probability, and our goal was to demonstrate a purification

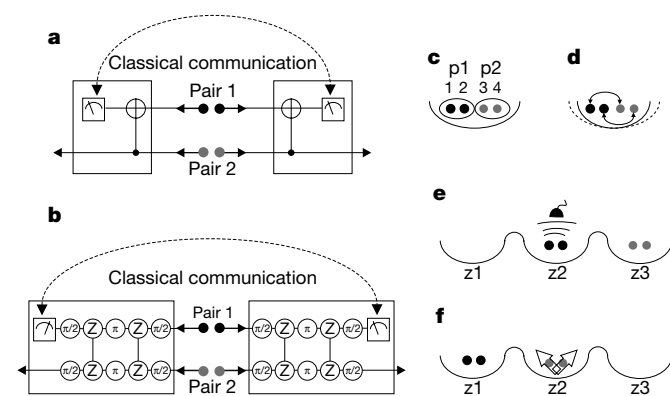


Figure 1 | Network diagrams for purification. **a**, The original proposal⁴. The qubits of both pairs are connected by CNOT gates across the pairs, after which the two qubits of pair 1 are measured. The measurement outcomes are compared by classical communication, and purification succeeds if they are the same. **b**, Experimental implementation in this work. The ion qubits of both entangled pairs are (1) rotated by $R(\pi/2, \pi/4)$ (see Methods), (2) connected by two-ion $\varepsilon = \pi/4$ -phase gates across the pairs (marked with Z), (3) rotated by $R(\pi, 5\pi/4)$, (4) connected by two-ion $\varepsilon = \pi/4$ -phase gates again and (5) finally rotated by $R(\pi/2, \pi/4)$. The ions of pair 1 are then measured, and purification succeeds if the two measurement outcomes are different. **c–f**, Gates applied to ions and positions of the ions. The potential well zones are shown schematically under the ions. **c**, With all ions trapped in one potential well, entangled pair p1 (black) and pair p2 (grey) are created by a phase gate as described in Methods. **d**, Corresponding ions in each pair are connected by phase gates (as shown in **b**). **e**, The pairs are separated into zones z2 and z3, and p1 is measured by state-dependent fluorescence. **f**, After moving the ions, the fidelity of p2 is determined in z2 by state tomography.

¹National Institute of Standards and Technology, Boulder, Colorado 80305, USA. †Present address: University of Ulm, 89069 Ulm, Germany.

protocol that is in principle useful for any sufficiently entangled input state. By randomizing the data for different input states, we determined the success of the protocol on effectively mixed states, a technique also used in ref. 6. We deterministically produced entangled pairs of $^9\text{Be}^+$ atomic ion qubits that were then distilled. The protocol succeeded with probabilities between 35% and 65%, depending on the initial fidelity. We demonstrated a gain in fidelity for a range of initial fidelities. In principle, the purified pairs are available for further rounds of the purification protocol or for use in other algorithms.

Our experimental procedure follows the original proposal of ref. 4 that is shown in Fig. 1a. We first confined four $^9\text{Be}^+$ ions in one trapping zone of a linear multi-zone Paul trap¹⁵ where they formed a linear array along the weakest axis of the trapping potential (Fig. 1c), which was approximately harmonic in three dimensions. Qubits were implemented with two hyperfine ground states of each ion, $|F=1, m_F=-1\rangle$ and $|F=2, m_F=-2\rangle$, designated respectively $|\uparrow\rangle$ and $|\downarrow\rangle$ for simplicity. We implemented qubit rotations $R(\theta, \phi)$ (see Methods) by driving stimulated Raman transitions with two laser fields¹⁵. Multi-qubit operations were realized with geometric phase gates generalized from the two-qubit phase gate described in ref. 16 (see Methods).

To prepare the initial state of the four-qubit register, all four modes of motion of the linear array were cooled to the ground state¹⁷, and the internal qubit states were optically pumped to $|\downarrow\rangle$. We then generated two pairs of qubits with varying degrees of entanglement, as depicted schematically in Fig. 1c. A phase gate embedded between rotations on all ions entangled pair 1 (p1, consisting of ions 1 and 2) and pair 2 (p2, consisting of ions 3 and 4) separately at the same time, while all four ions were held in the same trap potential well. Ideally, after this operation the state of the two pairs is:

$$|\Phi\rangle_{p1} \otimes |\Phi\rangle_{p2} = (\cos(\varepsilon)|\downarrow\downarrow\rangle_{12} + i\sin(\varepsilon)|\uparrow\uparrow\rangle_{12}) \otimes (\cos(\varepsilon)|\downarrow\downarrow\rangle_{34} + i\sin(\varepsilon)|\uparrow\uparrow\rangle_{34}) \quad (1)$$

For angle $\varepsilon = \pi/4$, the state is a product of two maximally entangled Bell pairs $|\Phi_{\text{Bell}}\rangle_j = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle_j + i|\uparrow\uparrow\rangle_j)$, with $j = \{p1, p2\}$. Experimentally, the states $|\Phi\rangle_j$ were obtained with a fidelity of approximately 0.75 for $\varepsilon = \pi/4$, limited by imperfections in the phase gate operation. Ideally, for general ε , the fidelity of each of the created pairs with respect to this Bell pair is:

$$F_{\text{pair}} = |{}_j\langle\Phi|\Phi_{\text{Bell}}\rangle_j|^2 = \cos^2(\varepsilon - \pi/4) \quad (2)$$

The states $|\Phi\rangle_j$ can be viewed as the ideal Bell pairs ($\varepsilon = \pi/4$) perturbed by a coherent combination of phase (sign flip) errors for $\varepsilon \neq \pi/4$. To lowest order in the deviation of ε from $\pi/4$, the errors act on just one of the qubits and are detectable by the purification process. By applying the purification protocol to this family of states, we are able to determine how well the experimental implementation performs on this type of phase error. To determine the performance on mixtures of these states, we use weighted combinations of the data for different ε .

We implemented a slight variation of the ref. 4 proposal shown in Fig. 1b. We used one purification phase gate connecting each member of the first entangled pair to its counterpart in the second pair (Fig. 1d). The first two qubits (p1) were then measured to determine whether purification succeeded (Fig. 1e). Execution of the phase gate ideally results in the state:

$$|\Psi\rangle = \frac{1}{2}(|\downarrow\uparrow\rangle_{12}|\downarrow\uparrow\rangle_{34} + |\uparrow\downarrow\rangle_{12}|\uparrow\downarrow\rangle_{34} + \cos(2\varepsilon)(|\downarrow\downarrow\rangle_{12}|\downarrow\downarrow\rangle_{34} + |\uparrow\uparrow\rangle_{12}|\uparrow\uparrow\rangle_{34}) + \sin(2\varepsilon)(|\downarrow\uparrow\rangle_{12}|\uparrow\downarrow\rangle_{34} + |\uparrow\downarrow\rangle_{12}|\downarrow\uparrow\rangle_{34})) \quad (3)$$

The quantum correlations between the entangled pairs are similar to those produced in the original proposal. However, measurement on

the first two qubits must show they are either $|\uparrow\downarrow\rangle$ or $|\downarrow\uparrow\rangle$ for the purification to succeed, and the resulting entangled state of qubits 3 and 4 is then rotated such that when $\varepsilon = \pi/4$ the state is $|\Psi_+\rangle \equiv \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$. For general ε , the probability of measuring that ions 1 and 2 are different is $P_s = \frac{1}{2}[1 + \sin^2(2\varepsilon)]$, in which case the state of the second pair of qubits is projected to:

$$|\Psi_{\text{purif}}\rangle = \frac{1}{\sqrt{1 + \sin^2(2\varepsilon)}} \times \begin{cases} (|\uparrow\downarrow\rangle + \sin(2\varepsilon)|\downarrow\uparrow\rangle) & \text{for first pair in } |\uparrow\downarrow\rangle \\ (|\downarrow\uparrow\rangle + \sin(2\varepsilon)|\uparrow\downarrow\rangle) & \text{for first pair in } |\downarrow\uparrow\rangle \end{cases} \quad (4)$$

Ideally, the purified fidelity is independent of whether $|\uparrow\downarrow\rangle$ or $|\downarrow\uparrow\rangle$ is detected and:

$$F_{\text{purif}} = |\langle\Psi_{\text{purif}}|\Psi_+\rangle|^2 = \frac{4\cos^4(\varepsilon - \pi/4)}{3 + \cos[4(\varepsilon - \pi/4)]} \geq F_{\text{pair}} \quad (0 \leq \varepsilon \leq \pi/2) \quad (5)$$

Near $\varepsilon = \pi/4$, the fidelity of the unpurified pairs decreases quadratically in $(\varepsilon - \pi/4)$, whereas the fidelity of the purified pair decreases quartically.

For the measurements and determination of the fidelity of the purified pair, we separated groups of ions in the multi-zone architecture of the trap¹⁵. We chose an order of the ions that made the measurement on p1 and the subsequent tomography on p2 as simple as possible. If we were distilling remote entanglement, then both pairs p1 and p2 would have one qubit in each location. Results of the

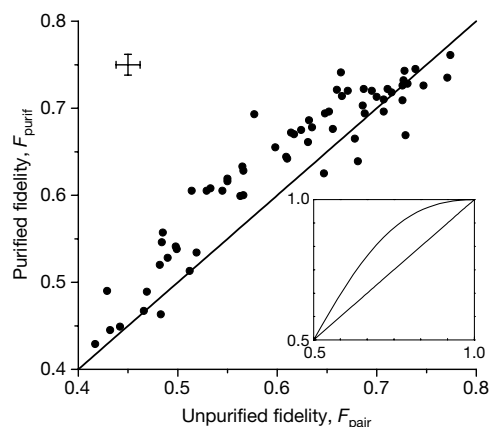


Figure 2 | Purified fidelity as a function of unpurified fidelity. The solid straight line in the main figure and inset represents $F_{\text{pair}} = F_{\text{purif}}$. For the experimental data points, F_{pair} was varied by experimentally setting angles $0 \leq \varepsilon \leq \pi/4$ to (ideally) obtain entangled pairs described by equation (1). Both fidelities were determined by (partial) state tomography on pair 2 as depicted in Fig. 1f, and are given by the probability that the pair was in state $|\Psi_+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$. For determining the unpurified fidelity, each purification experiment was immediately followed by an experiment that was identical except for the omission of the purifying gate, which are the operations inside the boxes of Fig. 1b. A rotation $R(\pi/2, -\pi/4)$, transforming $|\Phi_{\text{Bell}}\rangle$ into $|\Psi_+\rangle$, was applied to both pairs instead. The standard error of the points is approximately 0.012 in both variables (see representative error bars shown in the graph), as estimated by resampling (see Methods). The scatter in the points, caused by drifts in experimental conditions, significantly exceeds this standard error (see Methods). The curved solid line in the inset shows the theoretical purified fidelity as a function of the input pair fidelity for the demonstrated protocol and perfect operations as expressed by equations (2) and (5). Axis designations are the same as for the main plot. For a mixture of our data that approximates a uniform distribution of angles ε in equation (1) between $\varepsilon = 0$ and $\varepsilon = \pi/4$, we obtained an unpurified fidelity of 0.614 ± 0.0015 , and a purified fidelity of 0.629 ± 0.0015 . This implies a statistically significant improvement in fidelity of 0.015 ± 0.002 for this particular mixture.

measurement on p1 would then have to be communicated between the locations (as depicted in Fig. 1a) to compare the local outcomes.

All ions were transferred into a separation zone and separated into the two initial pairs within 380 μ s. Pair p1 returned to the original interaction zone, where the state of the two ions was detected by state-dependent fluorescence (Fig. 1e). We used a detection period of 200 μ s, during which we detected about 0.5 photons if both ions were in state $|\uparrow\rangle$, and about 10 photons on average for each ion in state $|\downarrow\rangle$. We declared the purification a success when the number of counts was in an interval determined by the property that the probability of falsely declaring success with both ions in state $|\uparrow\rangle$ ($|\downarrow\rangle$) was less than 0.01 (0.05). We then moved pair p1 out of the interaction zone and the purified pair p2 into it (Fig. 1f) in about 470 μ s. The state of p2 was analysed by a tomography procedure that determined all four Bell-state populations without distinguishing between the ions. This yielded an experimental value of the fidelity F_{purif} (see Methods). The complete experimental sequence, including tomography, took about 1.54 ms.

For comparison, each attempt at purification was followed by a reference experiment in which the entangled pairs were prepared in the same way but the purification phase gate was omitted. Instead we applied a common rotation $R(\pi/2, -\pi/4)$ to both pairs, transforming $|\Phi_{\text{Bell}}\rangle$ into $|\Psi_{+}\rangle$. The fidelity F_{pair} of pair p2 was then determined by the same tomography procedure. The fidelity of p1 was characterized in separate experiments and found to be equal to that of p2 within experimental uncertainties. Therefore F_{pair} is the fidelity of a prepared Bell pair if purification was not used. The same noise processes act on the ions in the purification and the reference experiment after the time in the procedure where the purification gate would be applied. To determine the effectiveness of purification, F_{pair} was compared to F_{purif} (Fig. 2). The final fidelity of the purified pair reached a maximum of about 0.75. Nevertheless, we demonstrated an improvement $F_{\text{purif}} > F_{\text{pair}}$ for $0.5 \leq F_{\text{pair}} \leq 0.7$. Ideally, neither fidelity should drop below 0.5 for the family of states used here. The apparent improvements discernible in the figure for $F_{\text{pair}} < 0.5$ are explained by the common sources of errors equally affecting the purification and reference experiments after the time when the purification gate is implemented (or not). Depending on the initial fidelity, the probability of success varied between 35% and 65%.

The data shown in Fig. 2 are for the family of states given in equation (1) perturbed by experimental noise. From this data, one can infer the performance of the purification protocol on mixtures of such states. We chose a mixture that approximates a uniform distribution of angles ε in equation (1) between $\varepsilon = 0$ and $\varepsilon = \pi/4$. The unpurified fidelity for this mixture was 0.614 ± 0.0015 , and the purified fidelity was 0.629 ± 0.0015 . This implies a statistically significant improvement in fidelity of 0.015 ± 0.002 for this particular mixture.

As the purification protocol took place with the two pairs in the same trapping zone and with operations simultaneously acting on all ions, it was necessary to verify that the implemented operations did not introduce unintended correlations. We confirmed experimentally that, subject to the limitations of the tomographic analysis, state preparation resulted in independent pairs of entangled ions and, if the purification gate was applied directly to the state $|\uparrow\uparrow\uparrow\uparrow\rangle$ or $|\downarrow\downarrow\downarrow\downarrow\rangle$, the state of ions 1 and 3 was independent of the state of ions 2 and 4, within experimental error.

Thresholds of tolerable error rates for entanglement purification in quantum repeaters are of the order of a few per cent, much less than the imperfections present in our current experiment. To improve state preparation and purification in future implementations, better control of classical parameters such as magnetic field and laser intensity will be required. Decoherence due to spontaneous emission could be reduced by an appropriate choice of the laser-beam detuning¹⁸. More advanced multi-zone traps and sympathetic cooling¹⁹ will make it possible to implement purification with the ions of each entangled pair transported to separated trap zones. With these improvements, purified Bell pairs with sufficient fidelity to violate

Bell inequalities should be feasible. Because the two pairs of entangled states in equation (1) have identical angles ε , our procedure tests the behaviour of the purification protocol in the case of collective phase errors. To determine the behaviour for other phase errors, one could use individual laser addressing of the ion or magnetic field gradients.

In summary, we have demonstrated entanglement purification with relatively high success rates in a potentially scalable system. The protocol and success rates demonstrated, together with the availability of the purified pair, could enable 'entanglement pumping' by repetitive application of the purification protocol. Ideally, the fidelity of the remaining pair(s) can be 'pumped' arbitrarily close to 1, but in practice will never exceed a limit imposed by imperfections in the purification gates and the measurements used to produce the purified pairs. In addition to uses in quantum communication and large-scale quantum information processing, remote entangled atoms could be useful in more fundamental experiments, such as a loop-hole-free test of local hidden-variable theories. The multi-segmented trap architecture used here should allow the distribution of entangled particles to separate locations for exploring repetitive protocols in future experiments.

METHODS

Gate operations. A general single-qubit rotation $R(\theta, \phi)$ transforms the qubit states as $R(\theta, \phi)|\uparrow\rangle = \cos(\theta/2)|\uparrow\rangle - ie^{i\phi}\sin(\theta/2)|\downarrow\rangle$, and $R(\theta, \phi)|\downarrow\rangle = -ie^{-i\phi}\sin(\theta/2)|\uparrow\rangle + \cos(\theta/2)|\downarrow\rangle$. These rotations were applied uniformly to all ions that resided in the trap zone z2 addressed by the Raman laser beams (Fig. 1c–f).

Phase gates for entangling different combinations of pairs of ions in a string of four ions are a generalization of the phase gate described in ref. 16. For the arrangement of laser beams in the experiment, the phase of the dipole force repeated every 213 nm along the alignment direction of the ions. The equilibrium between mutual Coulomb repulsion and the confinement of the external trap potential determined the positions of the ions relative to the trap centre to be $s(-1.437, -0.454, 0.454, 1.437)$, with $s = \sqrt{e^2/(4\pi\epsilon_0 m\omega_{\text{COM}}^2)}$, e the elementary charge, ϵ_0 the vacuum permittivity, m the mass of the beryllium ion and $\omega_{\text{COM}}/(2\pi)$ the axial centre of mass (COM) frequency²⁰. For ${}^9\text{Be}^+$, $s \approx 7.31 \mu\text{m}$ ($\omega_{\text{COM}}/2\pi[\text{MHz}]^{-2/3}$). By choosing the strength of the external potential appropriately, we achieved a pattern of dipole forces that equally coupled ion 1 to 2, and ion 3 to 4, with negligible coupling of other ion pairs. Within a coupled pair, phases changed according to $(|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle) \rightarrow (|\uparrow\uparrow\rangle, e^{i\varepsilon}|\uparrow\downarrow\rangle, e^{i\varepsilon}|\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle)$, where ε was determined by the time of the operation and the beam intensities¹⁶. To entangle two pairs of ions, we first applied a global rotation $R(\pi/2, 0)$ to all four ions, followed by the phase gate with $0 \leq \varepsilon \leq \pi/4$ on the axial mode at $\omega = 2\pi \times 6.742 \text{ MHz} \approx 2.41\omega_{\text{COM}}$, which has normalized mode amplitudes $(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$. We then applied a refocussing $R(\pi, \pi)$ rotation, the same phase gate, and finally another $R(\pi/2, 0)$ rotation. Pair preparation took less than 100 μ s. Calculations show that residual phases due to unwanted couplings only degrade the pair fidelity by 0.007. The observed loss in fidelity can be attributed to the following causes: during the pair preparation procedure, there was a 7% probability of one of the ions undergoing an absorption–spontaneous emission cycle. Phase instabilities of the single qubit rotations caused by interferometric phase fluctuations between Raman beams contributed of the order of 10% over the 1.5-ms duration of the experiment. Further imperfections included laser beam pointing and intensity noise, leading to fluctuations in the ion–laser couplings (Rabi frequencies) of the order of 5%. We also estimated an error of 5% due to fluctuations in the trapping potential, caused by noisy potentials applied to the trap electrodes. The last three errors also included systematic drifts that contributed to the statistically significant scatter in the measured fidelities (Fig. 2).

The purification gate was similar to the first gate. Two phase gates Z with angle $\varepsilon = \pi/4$ were embedded into single qubit rotations in a sequence $R(\pi/2, \pi/4) - Z - R(\pi, 5\pi/4) - Z - R(\pi/2, \pi/4)$ applied to all ions. The phase gates Z coupled ion 1 to 3 and 2 to 4 (see Fig. 1c) and were executed on the COM mode at a frequency of 4.07 MHz (duration 65 μ s). Calculations show that unwanted couplings in this gate degrade the fidelity by 0.007. The actual loss of fidelity due to the purification procedure was similar to those discussed above. Additional loss of fidelity in the purified pair was caused by imperfect state discrimination of pair 1 (see main text). The photon count thresholds for inferring that exactly one ion was in state $|\uparrow\rangle$ were set so that the probability of error is at most 5%.

State tomography. During detection (duration 200 μ s) we registered between $0.15 \leq \lambda_0 \leq 0.8$ counts if all ions are projected into $|\uparrow\rangle$, and between $8 \leq \lambda_1 \leq 12$ additional average counts for each ion in state $|\downarrow\rangle$. Count averages λ_0 and λ_1 were derived by fitting mixtures of poissonian distributions to count histograms for

the relevant detection periods and to reference histograms obtained by preparing all ions to be observed in the $|\downarrow\rangle$ or $|\uparrow\rangle$ states. We used the maximum likelihood method for fitting the histograms, and parametric-bootstrap resampling for determining standard errors in inferred quantities²¹. To obtain the fidelity of the second Bell pair (ions 3 and 4), we applied the tomographic rotation $T(\phi_a, \phi_b) = R(\pi/2, \phi_a) R(\pi/2, \phi_b)$ to both ions simultaneously, with $\phi_a = 0, \pi/4, \pi/2$ and $\phi_b = 0, \pi/4, \dots, 7\pi/4$. From the photon counts obtained after the tomographic rotations, we inferred entries of the density matrix in the Bell basis before the rotations by maximum likelihood methods. The inferred entries included the Bell state populations but not the coherences between the singlet and triplet states. We used a version of the maximum likelihood algorithm described in ref. 22. In the reference experiment, we extended the tomographic inference by including the counts for the measurement of ions 1 and 2, which yielded the symmetrized density matrices on ions 3 and 4 conditional on the number of ions 1 and 2 in state $|\uparrow\rangle$. We used this information to check that the state in the reference experiment was consistent with the first and second pair being independently entangled. In separate experiments, we checked that the behaviour of the purification gate was consistent with correlating only ions 1 and 3, and (separately) ions 2 and 4. In both cases, we found the inferred density matrix to be consistent with the independence assumption, in the sense that they match the terms of an independent density matrix with probability of error less than 0.01.

Received 19 June; accepted 3 August 2006.

- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Bennett, C. H. *et al.* Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
- Kwiat, P. G., Barraza-Lopez, S., Stefanov, A. & Gisin, N. Experimental entanglement distillation and 'hidden' non-locality. *Nature* **409**, 1014–1017 (2001).
- Pan, J.-W., Gasparoni, S., Ursin, R., Weihs, G. & Zeilinger, A. Experimental entanglement purification of arbitrary unknown states. *Nature* **423**, 417–422 (2003).
- Yamamoto, T., Koashi, M., Özdemir, S. K. & Imoto, N. Experimental extraction of an entangled photon pair from two identically decohered pairs. *Nature* **421**, 343–346 (2003).
- Zhao, Z., Yang, T., Chen, Y.-A., Zhang, A.-N. & Pan, J.-W. Experimental realization of entanglement concentration and a quantum repeater. *Phys. Rev. Lett.* **90**, 207901 (2003).
- Walther, P. *et al.* Quantum nonlocality obtained from local states by entanglement purification. *Phys. Rev. Lett.* **94**, 040504 (2005).
- Quantum Information Science and Technology Experts Panel. ARDA quantum information science and technology roadmap. (<http://qist.lanl.gov>) (2002).
- Browne, D. E. & Rudolph, T. Resource-efficient linear optical quantum computation. *Phys. Rev. Lett.* **95**, 010501 (2005).
- Eisert, J., Jacobs, K., Papadopoulos, P. & Plenio, M. B. Optimal local implementation of nonlocal quantum gates. *Phys. Rev. A* **62**, 052317 (2000).
- Brennen, G. K., Song, D. & Williams, C. J. Quantum computer architecture using nonlocal interactions. *Phys. Rev. A* **67**, 050302 (2003).
- Knill, E. Quantum computing with realistically noisy devices. *Nature* **434**, 39–44 (2005).
- Barrett, M. D. *et al.* Deterministic quantum teleportation of atomic qubits. *Nature* **429**, 737–739 (2004).
- Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
- King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
- Ozeri, R. *et al.* Hyperfine coherence in the presence of spontaneous photon scattering. *Phys. Rev. Lett.* **95**, 030403 (2005).
- Barrett, M. D. *et al.* Sympathetic cooling of $^9\text{Be}^+$ and $^{24}\text{Mg}^+$ for quantum logic. *Phys. Rev. A* **68**, 042302 (2003).
- Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Nat. Inst. Stand. Technol.* **103**, 259–328 (1998).
- Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
- Lvovsky, A. I. & Raymer, M. G. Continuous-variable optical quantum state tomography. Preprint at (<http://www.arXiv.org/quant-ph/0511044>) (2005).

Acknowledgements This work was supported by the Disruptive Technology Office (DTO), by a DoD Multidisciplinary University Research Initiative (MURI) programme administered by the Office of Naval Research and by NIST. R.R. was supported by the Alexander von Humboldt Foundation. We thank S. Glancy and W. Itano for comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L. (dil@boulder.nist.gov).

LETTERS

High-throughput synthesis and catalytic properties of a molecular sieve with 18- and 10-member rings

Avelino Corma¹, María J. Díaz-Cabañas¹, José Luis Jordá¹, Cristina Martínez¹ & Manuel Moliner¹

Crystalline molecular sieves with large pores and high adsorption capacities have many potential applications^{1–4}. Of these materials, zeolites are of particular interest owing to their stability in a wide range of experimental conditions. An aluminophosphate with very large circular channels⁵ containing 18 oxygen atoms (18-ring channels) has been synthesized, but in the search for large-pore zeolites, most of the materials which have been synthesized up to now contain only 14-ring channels; the synthesis of zeolites with larger ring structures has been believed to be hindered by the low Si–O–Si bond angles available. A silicogalioaluminate (ECR-34) with unidirectional 18-ring channels¹² was recently reported, but exhibited low micropore volume, thus rendering the material less attractive for catalytic applications. Here we report the structure and catalytic activity of the silicogermanate zeolite ITQ-33; this material exhibits straight large pore channels with circular openings of 18-rings along the *c* axis interconnected by a bidirectional system of 10-ring channels, yielding a structure with very large micropore volume. The conditions for synthesis are easily accessible, but are not typical, and were identified using high-throughput techniques.

In recent work we showed that by combining the directing effects of fluoride ions and germanium framework atoms, we could synthesize new high-pore-volume zeolitic topologies containing a large number of double four-ring (D4R) units in the structure^{10,13,14}. This was one step towards the synthesis of extra-large high-pore-volume zeolites, since Brunner and Maier¹⁵ predicted that these structures should contain a relatively large number of 3- and 4-rings.

The silicogermanate ITQ-33 also contains 3- and 4-rings in the structure and can be synthesized in either OH[−] or F[−] media using hexamethonium, which is relatively inexpensive, as the structure-directing agent. We selected an initial experimental factorial design (3 × 4³), in which Si/Ge, T^{III}/(Si+Ge), OH[−]/(Si+Ge), and H₂O/(Si+Ge) are the synthesis variables (see Supplementary Fig. 1). The alkali synthesis conditions for producing ITQ-33 are very unusual, owing to the low concentration of mobilizing agent required (OH[−]/(Si+Ge) = 0.1). The material is also synthesized,

with high crystallinity, in fluoride media (Fig. 1) with Al as trivalent and Si and Ge as tetravalent cations, forming crystals of ~0.2 μm (Supplementary Fig. 2). The chemical and thermogravimetric analyses of the as-prepared ITQ-33 indicate a Si/Ge/Al/N/F molar ratio of 1/0.45/0.06/0.2/0.1 and a total weight loss of 23%, consistent with the chemical formula [(hexamethonium)_{0.07}F_{0.07}(H₂O)_{0.37}][Si_{0.66}Al_{0.04}Ge_{0.30}O_{2.02}].

The X-ray diffraction (XRD) pattern analysis of calcined ITQ-33 is shown in Fig. 2. This zeolite possesses space group symmetry *P6/mmm*, with *a* = 19.367 Å and *c* = 11.495 Å (see Methods for details). The structure of ITQ-33 comprises a tertiary building unit formed by a [3²4³6⁹] cage with two additional atoms inside it, located along the direction of the *c* axis. This tertiary building block is formed by clustering three [45²6²] and two [35³] secondary building units, which share the T1–T3 edges, as shown in Fig. 3a. The presence in the structure of the 3-ring units is unusual^{11,16–19}.

The tertiary building units of ITQ-33 give rise to the formation of columns along the *c* axis by sharing the 3-ring with the neighbouring units (Fig. 3b). Each column is linked to three neighbour columns through the 4-rings, forming D4R cages (Fig. 3c), yielding two clusters and three D4R cages per unit cell. The ITQ-33 structure contains four symmetrically independent T-sites (Fig. 3a), and Rietveld refinement of the XRD pattern indicates the preferential isomorphic replacement of Si atoms with Ge in position T4 (corresponding to D4R) with 39% Ge, and T1 (position directly linked to the D4R) with 33% Ge; the other two positions remain as pure Si or (Si+Al). A similar preferential occupation of Ge in D4R has been reported for other zeolites^{9,20–22}. The chemical composition obtained from the refinement is [(Si+Al)_{0.71}Ge_{0.29}O₂], which closely matches that obtained by chemical analysis of the calcined sample ([H_{0.04}][Si_{0.66}Al_{0.04}Ge_{0.30}O₂]).

The topology of ITQ-33 is that of a 3-directional channel system in which there are straight extra-large channels with circular openings of 18 T-atoms along the *c* axis direction (Fig. 3e), leading to a crystallographic pore diameter of 12.2 Å. Up to now, circular pore openings bigger than 10 Å have only been observed in ECR-34 (10.1 Å)¹²

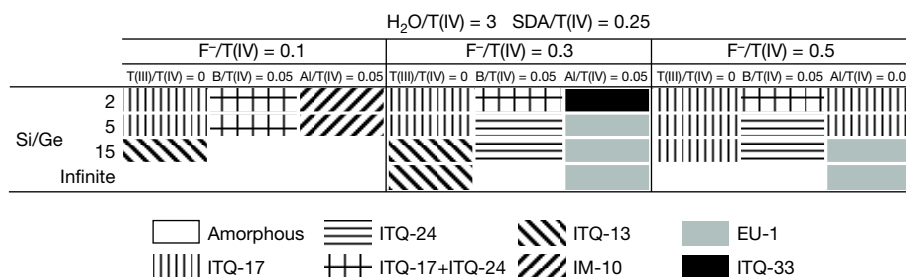


Figure 1 | Phase diagram in F[−] media.

¹Instituto de Tecnología Química, UPV-CSIC, Universidad Politécnica de Valencia, Avda. de los Naranjos s/n, 46022 Valencia, Spain.

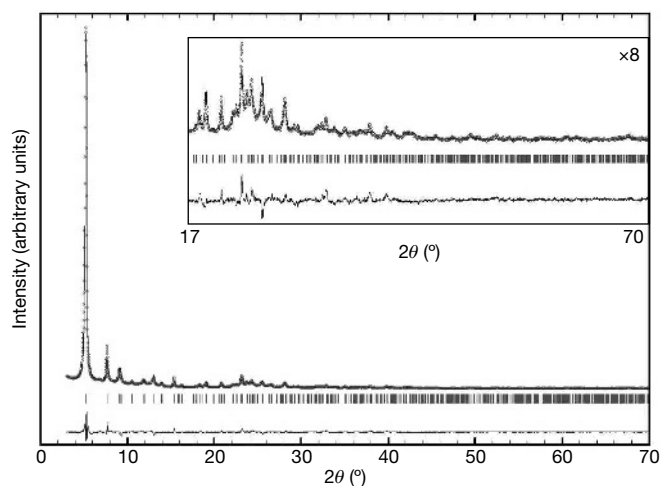


Figure 2 | Rietveld refinement of the XRD pattern of ITQ-33 calcined at 823 K. Observed (circles) and calculated XRD patterns, as well as the difference profile (bottom) are shown. The short tick marks below the pattern give the positions of the Bragg reflections. To show more detail, the scale for the pattern in the 2θ range of $17\text{--}70^\circ$ has been increased by a factor of eight in the inset.

and the aluminophosphate VPI-5 ($12.7\text{ \AA}$)⁵. However, an important difference between ITQ-33 and ECR-34 or VPI-5 is the presence in the former of a system of 10-ring channels in the plane $a\text{--}b$, with openings of $6.1 \times 4.3\text{ \AA}$, which interconnect the 18-ring channels (Fig. 3d). This gives a framework density for ITQ-33 of 12.3 T-atoms per $1,000\text{ \AA}^3$, which is much lower than for ECR-34 ($15.4\text{ T-atoms}/1,000\text{ \AA}^3$) and VPI-5 ($14.5\text{ T-atoms}/1,000\text{ \AA}^3$), both with 18-ring pores.

The structure-directing agent of ITQ-33 can be removed by calcination at 823 K, keeping the structure intact. We note that, owing to the size of the structure-directing agent, which is smaller than the zeolite pore diameter, 25% of the hexamethonium could be removed by ion exchange with NH_4Cl .

The N_2 adsorption properties of ITQ-33, considered per TO_2 , are much higher than that reported for any zeolite, but are still lower than those for MCM-41 (Supplementary Fig. 3). However, ITQ-33 starts to fill the gap between conventional zeolites and mesoporous

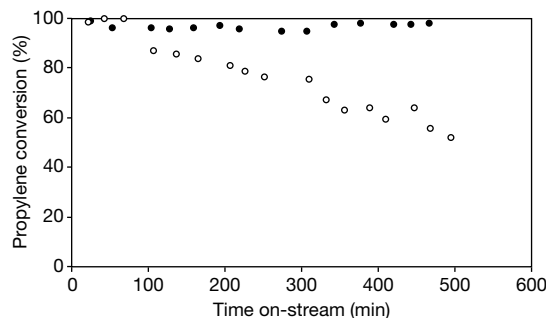


Figure 4 | Percentage propylene conversion as a function of time on-stream for ITQ-33 and BETA. ITQ-33, filled circles; BETA, open circles. $T = 125^\circ\text{C}$; $P = 3.5\text{ MPa}$; $\text{WHSV} = 12\text{ h}^{-1}$; the benzene/propylene molar ratio is 3.5.

materials. The Brunauer–Emmett–Teller (BET) surface area is $690\text{ m}^2\text{ g}^{-1}$ and the micropore volume $0.30\text{ cm}^3\text{ g}^{-1}$. We note that the estimated textural properties for the pure silica ITQ-33 would be $0.37\text{ cm}^3\text{ g}^{-1}$ of micropore volume and $840\text{ m}^2\text{ g}^{-1}$ of BET surface area.

The ^{27}Al magic angle spinning (MAS) nuclear magnetic resonance (NMR) spectrum of as-made ITQ-33 sample, shown in Supplementary Fig. 4, consists of a single resonance at 53 p.p.m., corresponding to tetrahedrally coordinated Al. No octahedral Al is detected. However, in the calcined sample a small amount of Al is extracted, as shown by a resonance at 0 p.p.m. The presence of tetrahedrally coordinated Al generates Brönsted acidity in the structure that can protonate pyridine, showing the characteristic band at $\sim 1,550\text{ cm}^{-1}$ in the infrared spectrum, characteristic of pyridinium ions (Supplementary Fig. 5). The presence of surface acidity and the connected extra-large pores in ITQ-33 also opens up new possibilities for catalysis, and will encourage further synthesis work to produce Ge-free Al-ITQ-33, as has recently been achieved with ITQ-24 (ref. 23).

Nevertheless, we can say that the surface acidity of the ITQ-33 sample is able to catalyse, with very high activity, interesting catalytic reactions such as alkylation of benzene with propylene to produce the industrially relevant cumene²⁴, while giving an extremely low yield of the undesirable n -propylbenzene (below 0.01% yield at 99% conversion). When working at very low contact time (weight hour space

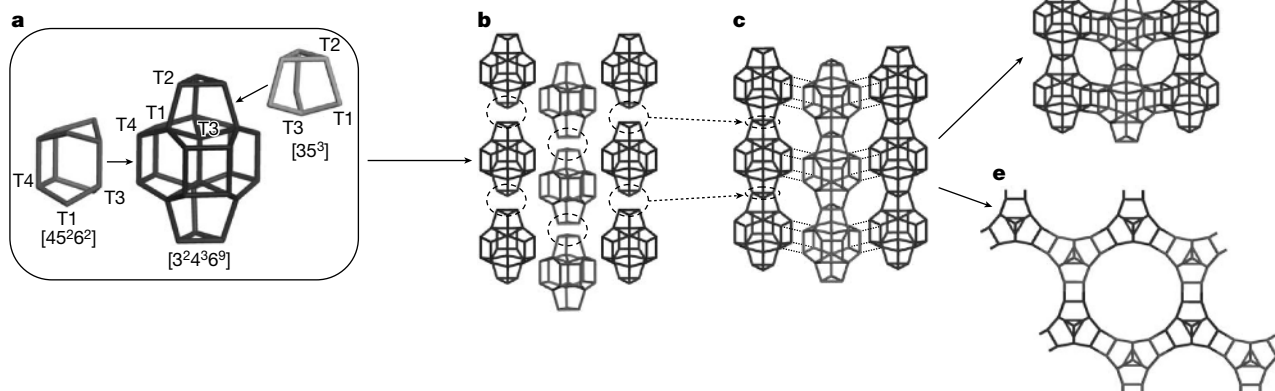


Figure 3 | Description of the structure of ITQ-33. **a**, The basic cluster of ITQ-33, and its constituent building units, indicating the location of the four independent T-atoms; **b**, condensation of the tertiary units in columns through the 3-ring; **c**, connection of the columns by formation of D4R cages

(dotted lines); **d**, lateral view of the ITQ-33 structure, showing the 10-ring; **e**, view of the complete structure along $[001]$, showing the 18-ring. Oxygen atoms have been removed for clarity.

Table 1 | Catalytic cracking of arabian light vacuum gasoil at 500 °C and 60 s time on-stream

Catalyst	Conversion (%)	Yields (%)			Molar ratio	
		Diesel	Gasoline	Propylene	Propylene/propane	Isobutene/isobutane
USY						
Catalyst/oil=0.62	92.5	15.7	40.4	4.7	1.0	0.1
Catalyst/oil=0.47	88.3	19.5	39.5	4.4	1.3	0.1
ITQ-33						
Catalyst/oil=0.70	89.2	22.6	34.5	4.2	1.9	0.4
Beta						
Catalyst/oil=0.70	84.0	14.1	32.3	7.5	1.9	0.5
USY + 20% ZSM-5	87.0	17.0	33.2	7.2	1.5	0.3
ITQ-33 + 20% ZSM-5	86.1	23.3	25.1	9.0	3.7	1.1

For the USY catalyst, unit cell size is 2.432 nm. The higher USY catalyst-to-oil ratio is close to the one used for ITQ-33 and Beta, and gives an idea about catalyst activity. The lower one allows us to achieve a level of conversion similar to other catalysts; this one is much better for comparing selectivities. The conversion is to diesel+gasoline+gases+coke. For diesel the boiling point is 216.1–359.0 °C and for gasoline it is 36.0–216.1 °C.

velocity, WHSV = 12 h⁻¹), ITQ-33 decays much more slowly than the Beta zeolite at present commercially used (Fig. 4). Activity is maintained after at least five reaction–regeneration cycles (540 °C in air). The material is very active for alkylation and transalkylation of alkyl aromatics, as well as for dealkylation of bulky alkylaromatics containing one or two condensed aromatic rings.

Finally, the reported structure should be of interest, compared with the existing ones, for producing more diesel and less gasoline, while maintaining the propylene and butene yield, during catalytic cracking of vacuum gasoil. This issue is of importance given that diesel has a higher mileage efficiency than gasoline and the necessity to save fuel and restrict CO₂ emissions. The catalytic behaviour predicted above would be a consequence of the very large pores that should increase the diesel yield, and the existence of the 10-ring connecting pores that allow diffusion and cracking of gasoline molecules, producing C₃ and C₄ olefins. The results presented in Table 1 indicate that ITQ-33 gives a cracking conversion higher than Beta, and close to that of a USY zeolite. Furthermore, ITQ-33 produces more diesel, less gasoline and propylene/propane and isobutene/isobutane ratios much higher than USY, and very similar to that of Beta zeolite. Used together, ITQ-33 and ZSM-5 have an excellent cooperative effect, giving a much higher diesel and propylene yield, with lower gasoline, than does the combination of USY and ZSM-5 zeolites used today. Zeolite ITQ-33 may thus transform the field of zeolite catalysis, if its stability and economics can be further improved.

METHODS

Synthesis of samples. Synthesis gels were prepared using an automatic system²³. The synthesis gel was introduced inside Teflon vials (3 ml), which were finally inserted in a 15-well multiautoclave. Crystallization was carried out at 175 °C under static conditions and the samples were characterized by XRD. Synthesis conditions are shown in Supplementary Methods.

Structure solution and refinement. The XRD pattern of a calcined ITQ-33 sample was collected as described in the Supplementary Methods. The pattern was indexed in a hexagonal unit cell with $a = 19.367$ Å and $c = 11.495$ Å, and the study of systematic extinctions suggested the space group $P6/mmm$ (no. 191).

The crystal structure was solved using the program FOCUS^{25,26}. Then, a Rietveld refinement was performed using the program FULLPROF²⁷, with a visually estimated background and a pseudo-Voigt profile function. The conditions of the refinement are shown in Supplementary Table 1. The residuals of the refinement were: $R_{\text{exp}} = 0.050$, $R_{\text{wp}} = 0.143$, $R_{\text{p}} = 0.046$, $R_{\text{B}} = 0.062$. Atomic coordinates, site occupancies and thermal parameters of the refined structure are shown in Supplementary Table 2. The agreement between observed and calculated patterns is shown in Fig. 2. The pure silica polymorph of this material was predicted by Treacy *et al.*²⁸, and can be seen in their database of hypothetical zeolite structures (<http://www.hypotheticalzeolites.net/>, reference number 191_4_19370.)

Catalytic tests. Alkylation of benzene with propylene was carried out in an automated high-pressure stainless steel reactor, at 125 °C, 3.5 MPa, with a benzene/propylene molar ratio of 3.5 and a WHSV = 12 h⁻¹. Catalytic cracking of a vacuum gasoil has been carried out in an automated microactivity test unit (ASTM D-3907) as described previously^{29,30} at 500 °C and 60 s time on-stream,

and at different catalyst-to-oil ratios (weight of zeolite/weight of gasoil fed in over 60 s). More details are given in the Supplementary Methods.

Received 1 June; accepted 7 September 2006.

- Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
- Jäger, R., Schneider, A. M., Behrens, P., Henkelmann, B., Schramm, K.-W. & Lenoir, D. Selective adsorption of polychlorinated dibenzo-p-dioxins and dibenzofurans by the zeolites UTD-1, SSZ-24, and ITQ-4. *Chem. Eur. J.* **10** (1), 247–256 (2004).
- Corma, A. State of the art and future challenges of zeolites as catalysts. *J. Catal.* **216**, 298–312 (2003).
- Ferey, G. Materials science. The simplicity of complexity-rational design of giant pores. *Science* **291**, 994–995 (2001).
- Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. A molecular sieve with eighteen-membered rings. *Nature* **331**, 698–699 (1988).
- Freyhardt, C. C., Tsapatsis, M., Lobo, R. F., Balkus, K. J. Jr & Davis, M. E. A high-silica zeolite with a 14-tetrahedral-atom pore opening. *Nature* **381**, 295–298 (1996).
- Yoshikawa, M. *et al.* Synthesis, characterization, and structure solution of CIT-5, a new, high-silica, extra-large-pore molecular sieve. *J. Phys. Chem. B* **102**, 7139–7147 (1998).
- Burton, A. *et al.* SSZ-53 and SSZ-59: two novel extra-large pore zeolites. *Chem. Eur. J.* **9** (23), 5737–5748 (2003).
- Paillaud, J. L., Harbuzaru, B., Patarin, J. & Bats, N. Extra-large-pore zeolites with two-dimensional channels formed by 14 and 12 rings. *Science* **304**, 990–992 (2004).
- Corma, A., Diaz-Cabanas, M. J., Rey, F., Nicolopoulos, S. & Boulahya, K. ITQ-15: the first ultralarge pore zeolite with a bi-directional pore system formed by intersecting 14- and 12-ring channels, and its catalytic implications. *Chem. Commun.* **12**, 1356–1357 (2004).
- Cheetham, T. *et al.* Very open microporous materials: from concept to reality. *Stud. Surf. Sci. Catal.* **135**, 788–795 (2001).
- Strohmaier, K. G. & Vaughan, D. E. Structure of the first silicate molecular sieve with 18-ring pore openings, ECR-34. *J. Am. Chem. Soc.* **125**, 16035–16039 (2003).
- Corma, A., Diaz-Cabanas, M. J., Martinez-Triguero, J., Rey, F. & Rius, J. A large-cavity zeolite with wide pore windows and potential as an oil refining catalyst. *Nature* **418**, 514–517 (2002).
- Corma, A., Navarro, M. T., Rey, F. & Valencia, S. Synthesis of pure polymorph C of Beta zeolite in a fluoride-free system. *Chem. Commun.* **16**, 1486–1487 (2001).
- Brunner, G. O. & Meier, W. M. Framework density distribution of zeolite-type tetrahedral nets. *Nature* **337**, 146–147 (1989).
- Annen, M. J., Davis, M. E., Higgins, J. B. & Schlenker, J. L. VPI-7: the first zincosilicate molecular sieve containing three-membered T-atom rings. *J. Chem. Soc. Chem. Commun.* **17**, 1175–1176 (1991).
- Zheng, N., Bu, X., Vu, H. & Feng, P. Open-framework chalcogenides as visible-light photocatalysts for hydrogen generation from water. *Angew. Chem. Int. Edn Engl.* **117**, 5433–5437 (2005).
- Correll, S., Oeckler, O., Stock, N. & Schnick, W. Li_xH_{12-x-y+z}P₁₂O_yN_{24-y}Cl_z—An oxonitridophosphate with a zeolite-like framework structure composed of 3-rings. *Angew. Chem. Int. Edn Engl.* **42**, 3549–3552 (2003).
- Lawton, S. L. & Rohrbach, W. J. The framework topology of ZSM-18, a novel zeolite containing rings of three (Si, Al)–O species. *Science* **247**, 1319–1322 (1990).
- Blasco, T. *et al.* Preferential location of Ge in the double four-membered ring units of ITQ-7 zeolite. *J. Phys. Chem. B* **106**, 2634–2642 (2002).
- Corma, A., Rey, F., Rius, J., Sabater, M. J. & Valencia, S. Supramolecular self-assembled molecules as organic directing agent for synthesis of zeolites. *Nature* **431**, 287–290 (2004).
- Wang, Y., Song, J. & Gies, H. The substitution of germanium for silicon in AST-type zeolite. *Solid State Sci.* **5**, 1421–1433 (2003).

23. Cantin, A., Corma, A., Diaz-Cabanas, M. J., Jorda, J. L. & Moliner, M. Rational design and HT techniques allow the synthesis of new IWR zeolite polymorphs. *J. Am. Chem. Soc.* **128**, 4216–4217 (2006).
24. Cejka, J. & Wichterlova, B. Acid-catalyzed synthesis of mono- and dialkyl benzenes over zeolites: active sites, zeolite topology, and reaction mechanisms. *Catal. Rev.* **44**, 375–421 (2002).
25. Grosse-Kunstleve, R. W., McCusker, L. B., Baerlocher, Ch. Powder diffraction data and crystal chemical information combined in an automated structure determination procedure for zeolites. *J. Appl. Cryst.* **30**, 985–995 (1997).
26. Grosse-Kunstleve, R. W., McCusker, L. B., Baerlocher, Ch. Zeolite structure determination from power diffraction data: applications of the FOCUS method. *J. Appl. Cryst.* **32**, 536–542 (1999).
27. Rodriguez-Carvajal, J. FULLPROF version 3.30, (<http://www-llb.cea.fr/fullweb/fp2k/fp2k.htm>) (Laboratory Leon Brillouin, 2005).
28. Treacy, M. M. J., Rivin, I., Balkowsky, E., Randall, K. H. & Foster, M. D. Enumeration of periodic tetrahedral frameworks. II. Polynodal graphs. *Micropor. Mesopor. Mater.* **74**, 121–132 (2004).
29. Corma, A. & Martinez-Triguero, J. The use of MCM-22 as a cracking zeolitic additive for FCC. *J. Catal.* **165**, 102–120 (1997).
30. Corma, A., Martinez-Triguero, J. & Martinez, C. The use of ITQ-7 as a FCC zeolitic additive. *J. Catal.* **197**, 151–159 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank CICYT for financial support, and Beamline BM25 (Spline) at ESRF for beam time allocation. J.L.J. thanks the Spanish MEC for a Juan de la Cierva contract and M.M. thanks CSIC for an I3P grant. We also thank F. Rey and V. Fornés for discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C. (acorma@itq.upv.es).

LETTERS

Eastern Pacific cooling and Atlantic overturning circulation during the last deglaciation

Markus Kienast¹, Stephanie S. Kienast¹, Stephen E. Calvert², Timothy I. Eglinton³, Gesine Mollenhauer⁴, Roger François² & Alan C. Mix⁵

Surface ocean conditions in the equatorial Pacific Ocean could hold the clue to whether millennial-scale global climate change during glacial times was initiated through tropical ocean–atmosphere feedbacks or by changes in the Atlantic thermohaline circulation¹. North Atlantic cold periods during Heinrich events and millennial-scale cold events (stadials) have been linked with climatic changes in the tropical Atlantic Ocean and South America^{2–4}, as well as the Indian and East Asian monsoon systems^{5,6}, but not with tropical Pacific sea surface temperatures⁷. Here we present a high-resolution record of sea surface temperatures in the eastern tropical Pacific derived from alkenone unsaturation measurements. Our data show a temperature drop of 1 °C, synchronous (within dating uncertainties) with the shutdown of the Atlantic meridional overturning circulation during Heinrich event 1, and a smaller temperature drop of 0.5 °C synchronous with the smaller reduction in the overturning circulation during the Younger Dryas event. Both cold events coincide with maxima in surface ocean productivity as inferred from ²³⁰Th-normalized carbon burial fluxes, suggesting increased upwelling at the time. From the concurrence of equatorial Pacific cooling with the two North Atlantic cold periods during deglaciation, we conclude that these millennial-scale climate changes were probably driven by a reorganization of the oceans' thermohaline circulation, although possibly amplified by tropical ocean–atmosphere interaction as suggested before⁸.

Heinrich events are anomalous occurrences of ice-rafted detritus in the North Atlantic between 70 and 10 kyr before present (BP; ref. 9), and there is growing consensus that these events were associated with substantial perturbations of the Atlantic meridional overturning circulation (AMOC)¹⁰. For example, based on the sedimentary ²³¹Pa/²³⁰Th ratio, a novel kinematic proxy for the vigour of the AMOC, it has been suggested¹⁰ that the meridional overturning was nearly, or completely, eliminated during Heinrich event 1 (H1).

Recent studies support a link between Heinrich events in the North Atlantic and changes in the environmental conditions of the tropical Atlantic Ocean and South America^{2,3}. Consistent with modelling studies¹¹, these palaeorecords suggest that a slowdown of the AMOC during H1 is accompanied by a southward displacement of the Intertropical Convergence Zone (ITCZ). Similarly, millennial-scale variability of the Indian and East Asian monsoon systems has been shown to be closely linked to climate change in the North Atlantic^{5,6}. The orchestrated response of the tropical Atlantic and the East Asian monsoon during the last glacial–interglacial transition has led to speculation¹² that the similarity in the deglacial evolution of climate in these distant regions is linked to shifts in the mean position

of the ITCZ. Whereas linkages between low and high northern latitude climate change are well established for the tropical Atlantic⁴ and the Indian⁵ and Asian⁶ monsoon systems, the interaction of the tropical Pacific Ocean with millennial-scale changes during the deglaciation observed elsewhere is still largely unknown. Thus, none of the few open oceanic Pacific sea surface temperature (SST) records available to date displays any obvious response to the reduction of the AMOC and/or the reorganization of atmospheric circulation during H1 and the Younger Dryas⁷.

Here we present a multi-proxy record of surface ocean conditions from the eastern equatorial Pacific (EEP) north of the Equator. Core ME0005A-24JC was recovered at 0° 1.3' N, 86° 27.8' W from 2,941 m water depth. According to the radiocarbon chronology (see Supplementary Information), sedimentation rates range between 23 and 29 cm kyr⁻¹, for the first time allowing resolution of centennial-scale changes in the EEP during the glacial–interglacial transition. Temperature estimates based on alkenone unsaturation (Supplementary Information) yield average values of 22.8 °C during the Last Glacial Maximum (19–23 kyr BP), about 2–3 °C below average latest Holocene (~25 °C from 2–4 kyr BP) or modern annual average (24.4–25.2 °C, World Ocean Atlas (WOA)) values (Fig. 1). The glacial–interglacial transition is punctuated by a cooling starting at ~18 kyr BP, with the coldest SST estimates (21.4–22.2 °C) between 16.5 and 14.5 kyr BP. Following this cooling event, alkenone unsaturation index $U_{37}^{K'}$ temperature estimates rise rapidly to 22.7–23.1 °C at ~14 kyr BP (Fig. 1). Within the acknowledged uncertainties of radiocarbon chronologies, the primary cold spell during the deglaciation is coeval with the H1 reduction of the AMOC (~17.5–14.5 kyr BP; ref. 10). The smooth rise of estimated SST to maximum values at the core surface is interrupted by a second cold event between 12.5 and 11.5 kyr BP, coeval with the Younger Dryas interval, with temperatures ~0.5 °C below the preceding and following time intervals (Fig. 1). Before the deglaciation (ages > 21 kyr BP), there is no obvious correlation to cooling events in the North Atlantic. However, uncertainties in our chronology before 21 kyr BP (Supplementary Information) preclude us from ruling out a linkage between the cold period in the EEP centred at 26.5 kyr BP (Fig. 1) and Heinrich events 2 (~24 kyr BP) or 3 (~29 kyr BP).

The higher sedimentation rates at this site are in part the result of sediment focusing (Supplementary Information), which has the potential to substantially bias alkenone-derived SST estimates¹³. It is therefore important to assess whether or not alkenone-derived EEP SST estimates presented here record local or regional surface ocean conditions or whether their variability is influenced by sediment advection and redeposition. Two sets of observations argue against a

¹Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, B3H 4J1 Canada. ²Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, V6T 1Z4 Canada. ³Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

⁴Alfred-Wegener Institute for Polar and Marine Research, 27570 Bremerhaven, Germany, and Department of Geosciences, University of Bremen, 28359 Bremen, Germany.

⁵College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA.

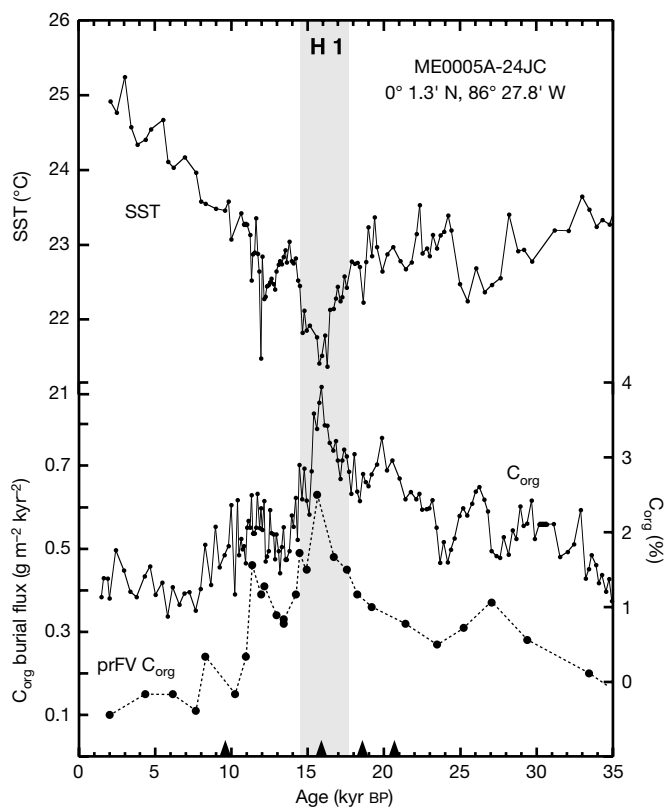


Figure 1 | Time series of data from site ME0005A-24JC in the eastern equatorial Pacific north of the Equator. Shown are estimated SST (based on alkenone unsaturation, U_{37}^K), percentage organic carbon (C_{org}) and ^{230}Th -normalized organic carbon burial flux (prFV C_{org}). The vertical bar marks the time period of the shut-down of the Atlantic meridional overturning circulation during Heinrich event 1¹⁰. Triangles on the bottom axis mark ^{14}C age control points.

significant sedimentological bias of the alkenone-based SST estimates. (1) Although ^{230}Th -normalization indicates significant sediment focusing during the Holocene (Supplementary Information), the average SST estimates of the topmost three samples ($\sim 25^\circ\text{C}$) agree very well with the annual mean SST at this site ($\sim 25.2^\circ\text{C}$, WOA). Likewise, the glacial–interglacial range of alkenone SST estimates presented here is in agreement with the range of glacial–interglacial SST variability obtained from lower-sedimentation-rate cores from this region; this latter range is based on Mg/Ca ratios (2.5°C at site TR163-19; ref. 14) and species assemblages (1 to 3°C at a number of EEP sites¹⁵) of planktonic foraminifera, which are less prone to redistribution by bottom currents than finer-grained sediments hosting the alkenones. Also, there is no correlation between periods of maximum focusing and minimum SSTs, or vice versa. (2) Detailed bathymetric maps and high-resolution chirp sub-bottom profiles^{16,17} show that core site ME0005A-24JC is located in a narrow east–west-trending trough in a region of abyssal hills, bordered by the Cocos–Nazca rise in the north and the Carnegie ridge to the south. Assuming that any potential sediment redistribution is topographically steered and bottom-current driven¹⁸, the source region for laterally advected sediments (including alkenones) is thus restricted to less than 100 km in the north–south direction. Within this general region, modern SSTs range through $\sim 2.3^\circ\text{C}$ (from 23.8°C 1.5° south of the Equator, to 26.1°C 1.5° north of the Equator, noting that WOA temperature gradients may be substantially smoothed relative to synoptic gradients), and glacial temperature gradients may have been either larger¹⁹ or smaller^{20,21}. To explain temperature changes of ~ 0.5 – 1.0°C exclusively in terms of sediment redistribution would require ~ 30 – 100% (depending on whether one adopts a typical synoptic gradient or the local WOA gradient, respectively) of the alkenone signal during the

cooling to have been reworked from the crest of the Carnegie ridge. Although the reworking hypothesis is not entirely impossible, we assume that the alkenone record presented here records local/meridional SST variability throughout the past 35 kyr despite evidence for substantial sediment focusing in this region (Supplementary Information).

Understanding the cause of the H1 cooling of the eastern tropical North Pacific will offer important insights into the mechanistic linkage between this region and the AMOC during this time interval. Two scenarios are entertained. A first mechanism that has been inferred to cause cooling of the EEP is increased advection of colder waters from the Peru Current^{22,23}. Two lines of reasoning, however, argue against advection of colder water masses as the principal driver of decreased SSTs in the EEP during H1. First, advection of colder waters to site ME0005A-24JC should also affect SSTs off the Galapagos Islands. However, foraminiferal Mg/Ca SST estimates from core V21-30 (ref. 21), which has high enough sedimentation rates to allow resolution of millennial-scale SST variability, do not show any cooling associated with this time interval. In addition, SST and palaeoproductivity records off Peru suggest an early warming of the Peru Current starting at ~ 20 kyr BP (ref. 22), without interruption during the H1 or Younger Dryas time intervals²⁴. A second possible explanation for cooling in the EEP north of the Equator is increased wind-driven upwelling (or upwelling of colder source waters). This scenario is supported by evidence for a close link between SST and marine primary production, as recorded by organic carbon percentage and ^{230}Th -normalized fluxes at our study site (Fig. 1).

The data presented here suggest that surface ocean conditions in the EEP varied synchronously (within the uncertainties of radiocarbon chronologies) with a slowdown of the AMOC during the time interval of H1 and the Younger Dryas. Two competing scenarios have been offered to explain abrupt climate changes, including H1 and the Younger Dryas event during the last deglaciation (see ref. 1 for a recent review). One attributes such changes to variations in the ocean's thermohaline circulation²⁵. The alternative view suggests that ocean–atmosphere interactions in the tropics, analogous to the present-day El Niño/Southern Oscillation (ENSO) dynamics, affect climate worldwide (see Supplementary Information, and ref. 26), possibly even through feedbacks on the AMOC⁸.

Perturbations to the ocean–atmosphere system of the tropical Pacific have the demonstrated potential to affect climate throughout the globe²⁷, suggesting that the tropical origin hypothesis has merit. However, a response of the tropical Pacific to orbital forcing as simulated by the Zebiak and Cane model cannot explain the presence of maximum productivity or minimum SSTs in the EEP coincident with both H1 and the Younger Dryas events²⁸. Although this highly idealized model shows that ENSO physics can abruptly lock to the seasonal cycle, thus leading to shutdowns of ENSO (that is, a mean state of the tropical Pacific resembling a La Niña event), the orbital configurations necessary for this to happen only recur on an approximately 11-kyr timescale, possibly synchronous with the Younger Dryas and Heinrich event 2, but not at the time of H1. On the other hand, newly developed global ocean–atmosphere models^{29,30} simulate significant changes in the tropical Pacific triggered by a southward displacement of the ITCZ associated with a slowdown of the AMOC, thus offering a number of potential (and non-exclusive) explanations for the observations presented here. In the models, a slowdown of the AMOC strengthens the northeast trade winds. Although the intensified trade winds north of the Equator could explain the decreased SSTs via latent cooling, the associated increase in mixing may not be enough to account for the significant increase in organic carbon flux at our site (Fig. 1). On the other hand, the close association of enhanced marine primary production with decreased SSTs could be indicative of increased equatorial upwelling driven by increased Ekman divergence caused by the intensified northeast trades. The inference of increased equatorial upwelling in the EEP,

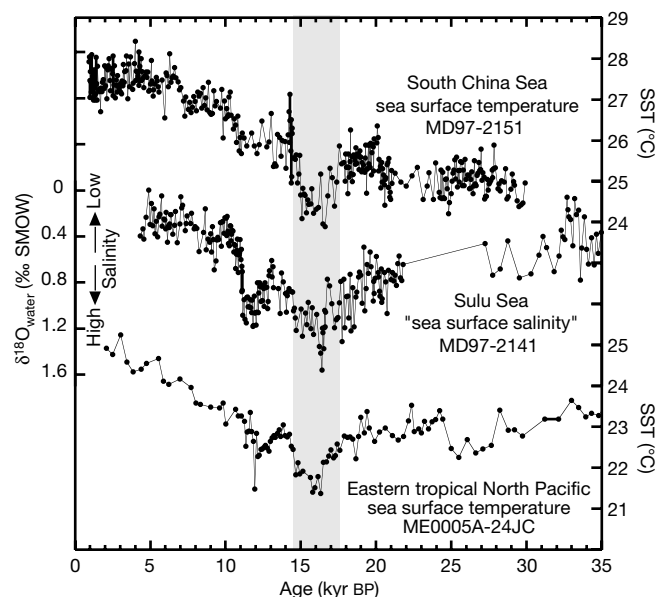


Figure 2 | Comparison of the SST evolution at site ME0005A-24JC to data from the Sulu Sea and the South China Sea. Shown are sea surface salinity ($\delta^{18}\text{O}_{\text{water}}$) estimates from the Sulu Sea³¹ and $U_{37}^{K'}$ SST estimates from the South China Sea³² on independent timescales. The vertical bar highlights climatic response to Heinrich event 1.

however, is not necessarily borne out by the recent coupled general circulation models^{29,30}, suggesting that the effect of intensified north-east trades on equatorial upwelling may be offset by weakened mean southeasterly cross-equatorial winds.

Finally, Zhang and Delworth²⁹ suggest that the intensified southward winds produced in their model in response to a shutdown of the AMOC induce coastal upwelling adjacent to central and south America (north of the Equator), leading to cooling of up to 0.5°C in the eastern tropical Pacific north of the Equator owing to propagation of the upwelling signal with Ekman drift. South of the Equator, the anomalous southward winds induce coastal downwelling, translating to a warming of $\sim 0.5^\circ\text{C}$ there. Our study site is essentially at the hinge point between these cool and warm anomalies in the model, which has a meridional resolution of the tropical ocean of $1/3^\circ$ and a latitudinal resolution of the atmosphere of 2° . These authors speculate²⁹ that a higher-resolution model could produce larger anomalies. For this model to explain the equatorial cooling and increased carbon flux during H1 and the Younger Dryas, we must assume that our site is dominated by the effects of the Northern Hemisphere coastal upwelling, and that our record from a single site is representative of the larger region of the EEP north of the Equator. This scenario could thus be confirmed by analysing additional high resolution cores in the region. The model predictions of Zhang and Delworth²⁹ are also in accord with lower SST and higher sea surface salinity reconstructions from the South China Sea and the Sulu Sea, respectively (Fig. 2), suggesting a close coupling of the position of the ITCZ, the tropical Pacific atmosphere–ocean circulation, and the East Asian monsoon (see also ref. 30).

Given the fact that the Zebiak and Cane model only accounts for changes due to variable insolation forcing, ignoring other forcings such as CO_2 or ice sheet extent, it might be premature to discount ENSO-related dynamics as the principal driver of abrupt climate change in the past. However, the records presented here lend support to the hypothesis that millennial-scale global climate change, at least during H1 and the Younger Dryas, is initiated by a reorganization of the ocean's thermohaline circulation, albeit potentially amplified by orbitally induced ENSO-related freshwater perturbations in the tropics (such amplification is considered in ref. 8).

Received 10 January; accepted 22 August 2006.

- Broecker, W. S. Does the trigger for abrupt climate change reside in the ocean or in the atmosphere? *Science* **300**, 1519–1522 (2003).
- Jenkinson, T. C. et al. Asynchronous terrestrial and marine signals of climate change during Heinrich events. *Science* **306**, 2236–2239 (2004).
- Wang, X. et al. Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature* **432**, 740–743 (2004).
- Peterson, L. C., Haug, G. H., Hughes, K. A. & Röhl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last glacial. *Science* **290**, 1947–1951 (2000).
- Schulz, H., von Rad, U. & Erlenkeuser, H. Correlation between Arabian Sea and Greenland climate oscillations of the past 110,000 years. *Nature* **393**, 54–57 (1998).
- Wang, Y. J. et al. A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
- Kiefer, T. & Kienast, M. Patterns of deglacial warming in the Pacific Ocean: a review with emphasis on the time interval of Heinrich event 1. *Quat. Sci. Rev.* **24**, 1063–1081 (2005).
- Schmittner, A. & Clement, A. C. Sensitivity of the thermohaline circulation to tropical and high latitude freshwater forcing during the last glacial-interglacial cycle. *Paleoceanography* **17**, doi:10.1029/2000PA000591 (2002).
- Hemming, S. R. Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42**, doi:10.1029/2003RG000128 (2004).
- McManus, J. F., François, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54**, 251–267 (2002).
- Lea, D. W., Pak, D. K., Peterson, L. C. & Hughes, K. A. Synchronicity of tropical and high-latitude Atlantic temperatures over the last glacial termination. *Science* **301**, 1361–1364 (2003).
- Ohkouchi, N., Eglinton, T. I., Keigwin, L. D. & Hayes, J. M. Spatial and temporal offsets between proxy records in a sediment drift. *Science* **298**, 1224–1227 (2002).
- Lea, D. W., Pak, D. K. & Spero, H. J. Climate impact of late Quaternary equatorial Pacific sea surface temperature variations. *Science* **289**, 1719–1724 (2000).
- Kucera, M. et al. Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans. *Quat. Sci. Rev.* **24**, 951–998 (2005).
- Lyle, M. et al. Do geochemical estimates of sediment focusing pass the sediment test in the equatorial Pacific? *Paleoceanography* **20**, doi:10.1029/2004PA001019 (2005).
- Mix, A. C., Tiedemann, R., Blum, P. & the Shipboard Scientific Party. Southeast Pacific Paleooceanographic Transects 1232–1242. *Proc. ODP Init. Rep.* **202**, (2003); (http://www-odp.tamu.edu/publications/202_IR/202ir.htm).
- Moore, T. C. Jr, Heath, G. R. & Kowman, R. O. Biogenic sediments of the Panama Basin. *J. Geol.* **81**, 458–472 (1973).
- Martinez, J. I., Keigwin, L. D., Barrows, T. T., Yokoyama, Y. & Southon, J. La Niña-like conditions in the eastern equatorial Pacific and a stronger Choco jet in the northern Andes during the last glaciation. *Paleoceanography* **18**, doi:10.1029/2002PA000877 (2003).
- Koutavas, A. & Lynch-Stieglitz, J. Glacial-interglacial dynamics of the eastern equatorial Pacific cold tongue-Intertropical Convergence Zone system reconstructed from oxygen isotope records. *Paleoceanography* **18**, doi:10.1029/2003PA000894 (2003).
- Koutavas, A., Lynch-Stieglitz, J., Marchitto, T. M. & Sachs, J. P. El Niño-like pattern in ice age tropical Pacific sea surface temperature. *Science* **297**, 226–230 (2002).
- Feldberg, M. J. & Mix, A. Planktonic foraminifera, sea surface temperatures, and mechanisms of oceanic change in the Peru and south equatorial currents, 0–150 ka BP. *Paleoceanography* **18**, doi:10.1029/2001PA000740 (2003).
- Lyle, M., Pahl, F. G. & Sparrow, M. A. Upwelling and productivity changes inferred from a temperature record in the central equatorial Pacific. *Nature* **355**, 812–815 (1992).
- Higginson, M. J. & Altabet, M. A. Initial test of the silicic acid leakage hypothesis using sedimentary biomarkers. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020511 (2004).
- Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
- McIntyre, A. & Molino, B. Forcing of Atlantic equatorial and subpolar millennial cycles by precession. *Science* **274**, 1867–1870 (1996).
- Cane, M. A. A role for the tropical Pacific. *Science* **282**, 59–61 (1998).
- Clement, A. C., Cane, M. A. & Seager, R. An orbitally driven tropical source for abrupt climate change. *J. Clim.* **14**, 2369–2375 (2001).
- Zhang, G. Y. & Delworth, T. L. Simulated tropical response to a substantial weakening of the Atlantic thermohaline circulation. *J. Clim.* **18**, 1853–1860 (2005).

30. Timmermann, A., Krebs, U., Justino, F., Goosse, H. & Ivanochko, T. S. Mechanisms for millennial-scale global synchronization during the last glacial period. *Paleoceanography* **20**, doi:10.1029/2004PA001090 (2005).
31. Rosenthal, Y., Oppo, D. W. & Linsley, B. K. The amplitude and phasing of climate change during the last deglaciation in the Sulu Sea, western equatorial Pacific. *Geophys. Res. Lett.* **30**, 1428–1431 (2003).
32. Huang, C.-C., Chen, M.-T., Lee, M.-Y., Wei, K.-Y. & Huang, C.-Y. Planktonic foraminifera faunal sea surface temperature records of the past two glacial terminations in the South China Sea near Wan-An shallow (IMAGES core MD97–2151). *West. Pacif. Earth Sci.* **2**, 1–14 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The manuscript benefited from discussions with A. Clement, T. Delworth, D. Lea and A. Timmermann. Laboratory assistance by M. Soon, D. Montluçon, A. Fleer and S. Brown-Leger is acknowledged. This work was supported by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), the Natural Sciences and Engineering Research Council (NSERC) of Canada, the National Science Foundation (NSF), the WHOI postdoctoral fellowship programme (M.K. and G.M.) and a fellowship of the Canadian Institute for Advanced Research (CIAR; M.K.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (markus.kienast@dal.ca).

LETTERS

Late survival of Neanderthals at the southernmost extreme of Europe

Clive Finlayson^{1,2}, Francisco Giles Pacheco³, Joaquín Rodríguez-Vidal⁴, Darren A. Fa¹, José María Gutierrez López⁵, Antonio Santiago Pérez³, Geraldine Finlayson¹, Ethel Allue⁶, Javier Baena Preysler⁷, Isabel Cáceres⁶, José S. Carrión⁸, Yolanda Fernández Jalvo⁹, Christopher P. Gleed-Owen¹⁰, Francisco J. Jimenez Espejo¹¹, Pilar López¹², José Antonio López Sáez¹³, José Antonio Riquelme Cantal¹⁴, Antonio Sánchez Marco⁹, Francisco Giles Guzman¹⁵, Kimberly Brown¹⁶, Noemí Fuentes⁸, Claire A. Valarino¹, Antonio Villalpando¹⁵, Christopher B. Stringer¹⁷, Francisca Martínez Ruiz¹¹ & Tatsuhiko Sakamoto¹⁸

The late survival of archaic hominin populations and their long contemporaneity with modern humans is now clear for southeast Asia¹. In Europe the extinction of the Neanderthals, firmly associated with Mousterian technology, has received much attention, and evidence of their survival after 35 kyr BP has recently been put in doubt². Here we present data, based on a high-resolution record of human occupation from Gorham's Cave, Gibraltar, that establish the survival of a population of Neanderthals to 28 kyr BP. These Neanderthals survived in the southernmost point of Europe, within a particular physiographic context, and are the last currently recorded anywhere. Our results show that the Neanderthals survived in isolated refuges well after the arrival of modern humans in Europe.

The association between Neanderthals and Gibraltar dates to 1848 when a Neanderthal cranium was discovered at Forbes's Quarry³. The excavation of the Devil's Tower rock shelter in 1925–26 by Dorothy Garrod produced a second Neanderthal cranium with associated Mousterian industry⁴. It has since been established beyond doubt that the Neanderthals were the makers of the Mousterian in western Europe⁵. In Gibraltar there are currently eight Neanderthal occupation sites on the 6-km-long 426-m-high rock. The presence of Mousterian technology in Gorham's Cave, Gibraltar, was established during excavations made between 1948 and 1954 (ref. 6). The site was revisited from 1995 with a view to establishing the timing of the Middle to Upper Palaeolithic transition. A date of 32.28 ± 0.42 kyr BP (OxA-7857) was obtained for the uppermost Mousterian levels⁷. Before 1997 all excavations and soundings had been made in the external part of the cave. Problems of contamination of radiocarbon samples from wet, unprotected, exterior, parts of caves have recently been brought to light². Here we describe the results of a series of excavations deep within Gorham's Cave between 1999 and 2005.

Excavations have been made of 29 m² of cave floor. The stratigraphy is composed of four levels (Fig. 1). Level I corresponds to Phoenician and Carthaginian (eighth to third century BC) horizons; level II corresponds to a brief occupation during the Neolithic. The

two levels of interest in this paper are levels III and IV. The technology associated with level III, from which 240 artefacts have been recovered so far, is Upper Palaeolithic. The raw materials are predominantly flints, cherts and some quartzites. Three horizons have been identified, two corresponding to occupation during the Solutrean (Supplementary Information) and a higher horizon that corresponds to the Magdalenian (Supplementary Information). Aurignacian and Gravettian are absent. The technology associated with level IV, with 103 artefacts recovered so far, is exclusively Mousterian with flints, cherts and quartzites (Supplementary Information). Previous excavations by Waechter⁶ also attributed the Upper Palaeolithic to the Solutrean and Magdalenian. Barton⁷ and Pettit & Bailey⁸ found only a few, undiagnostic, Upper Palaeolithic pieces in their profiles.

Figure 1 and Table 1 present 30 accelerator mass spectrometry (AMS) dates in stratigraphic position for levels III and IV in the deepest part of the cave. The samples were identified as individual pieces of charcoal under a microscope before being dated. The 22 AMS dates from the upper part of level IV span a 10-kyr period between 33 and 23 kyr BP. We have not attempted to calibrate the dates in view of the uncertainty surrounding calibration beyond 26 kyr cal BP⁹. The dates suggest a favoured location that was visited repeatedly over many thousands of years. Its situation, where natural light penetrates deep into the cave and where a high ceiling permits ventilation of smoke, is unique within the cave system, and hearths were made in the same location many times. This repeated use is confirmed by the stratigraphic distribution of the dates within level IV that indicate localized alterations due to use and reuse (for example trampling and cleaning) in the area around the position of the hearths but dates in stratigraphic sequence within the location of the hearths themselves. Thus, three samples (16, 17 and 20; Fig. 1) came from *in situ* Mousterian superimposed hearths. These three dates provide a stratigraphic sequence from $24,010 \pm 320$ to $30,560 \pm 720$ yr BP. Taken together, all the dates show that Neanderthals occupied the site until 28 kyr BP and possibly as recently as 24 kyr BP. The evidence in support of the 24 kyr BP date is more limited than for 28 kyr BP, which is taken as the latest well-supported

¹The Gibraltar Museum, 18–20 Bomb House Lane, Gibraltar. ²Department of Social Sciences, University of Toronto at Scarborough, Toronto, Ontario M1C 1A4, Canada. ³Museo Arqueológico de El Puerto Santa María, 11500 El Puerto Santa María, Spain. ⁴Departamento de Geodinámica y Paleontología, Facultad de Ciencias Experimentales, Universidad de Huelva, 21071 Huelva, Spain. ⁵Museo Municipal, 11650 Villamartin, Spain. ⁶Institut Català de Paleoeologia Humana i Evolució Social, Àrea de Prehistòria, Universitat Rovira i Virgili, 43003 Tarragona, Spain. ⁷Departamento de Prehistoria y Arqueología, Universidad Autónoma, 28049 Madrid, Spain. ⁸Department of Plant Biology, Universidad de Murcia, 30100 Murcia, Spain. ⁹Museo Nacional de Ciencias Naturales (CSIC), 28006 Madrid, Spain. ¹⁰The Herpetological Conservation Trust, Bournemouth, Dorset BH1 4AP, UK. ¹¹Instituto Andaluz de Ciencias de la Tierra, CSIC-UGR, 18002 Granada, Spain. ¹²Department of Geography, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, UK. ¹³Laboratorio de Arqueobotánica, Instituto de Historia (CSIC), 28014 Madrid, Spain. ¹⁴Departamento de Prehistoria y Arqueología, Universidad de Granada, 18071 Granada, Spain. ¹⁵Área de Prehistoria, Facultad de Filosofía y Letras, Universidad de Cádiz, 11003 Cádiz, Spain. ¹⁶Department of Biological Anthropology, University of Cambridge, Cambridge CB2 3DZ, UK. ¹⁷Department of Palaeontology, The Natural History Museum, London SW7 5BD, UK. ¹⁸Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, Yokosuka 237-0061, Japan.

occupation date. The level III dates represent a different occupation pattern and are consistent with sporadic visits to the cave during the Solutrean and Magdalenian and are separated from the Mousterian occupation by a stratigraphic gap of at least 5 kyr (Fig. 1 and Table 1).

Geochemical detrital ratios such as K/Al and Mg/Al are constant within each level at Gorham's Cave and different from each other. Level IV has a particular composition that differs from the rest of the sequence (Supplementary Information). Representative samples from all levels were analysed geochemically and mineralogically (Supplementary Information). The resulting data have been used as proxies to establish a sedimentary regime for the deep part of the cave, geochemical ratios acting as fingerprints in each level¹⁰. The sediments sampled were predominantly composed of clay minerals, calcite and quartz, with small quantities of dolomite, ankerite and feldspars. Levels III and IV are clearly differentiated. Level IV has a

geochemical characterization that is clearly different from other levels that contain Mg/Al and K/Al ratios that are almost twice as high (Supplementary Information). A sharp change in concentration of most of the elements analysed was detected between levels III and IV, suggesting a sudden change in environmental conditions and/or occupational pattern. Together with the absence of Upper Palaeolithic tools in level IV and of any dates older than 19 kyr BP in level III, the geochemical results confirm that there is no contamination of level IV from the upper levels and that the boundary between levels III and IV is sharp and clear.

Taphonomic study (Supplementary Information) indicated that all taxa and animal sizes, according to weight, showed evidence of human-induced damage (for example cut-marks, percussion scars and conchoidal breakage), affecting 14.4% of identified specimens. This is in agreement with other Middle Palaeolithic sites^{11,12}.

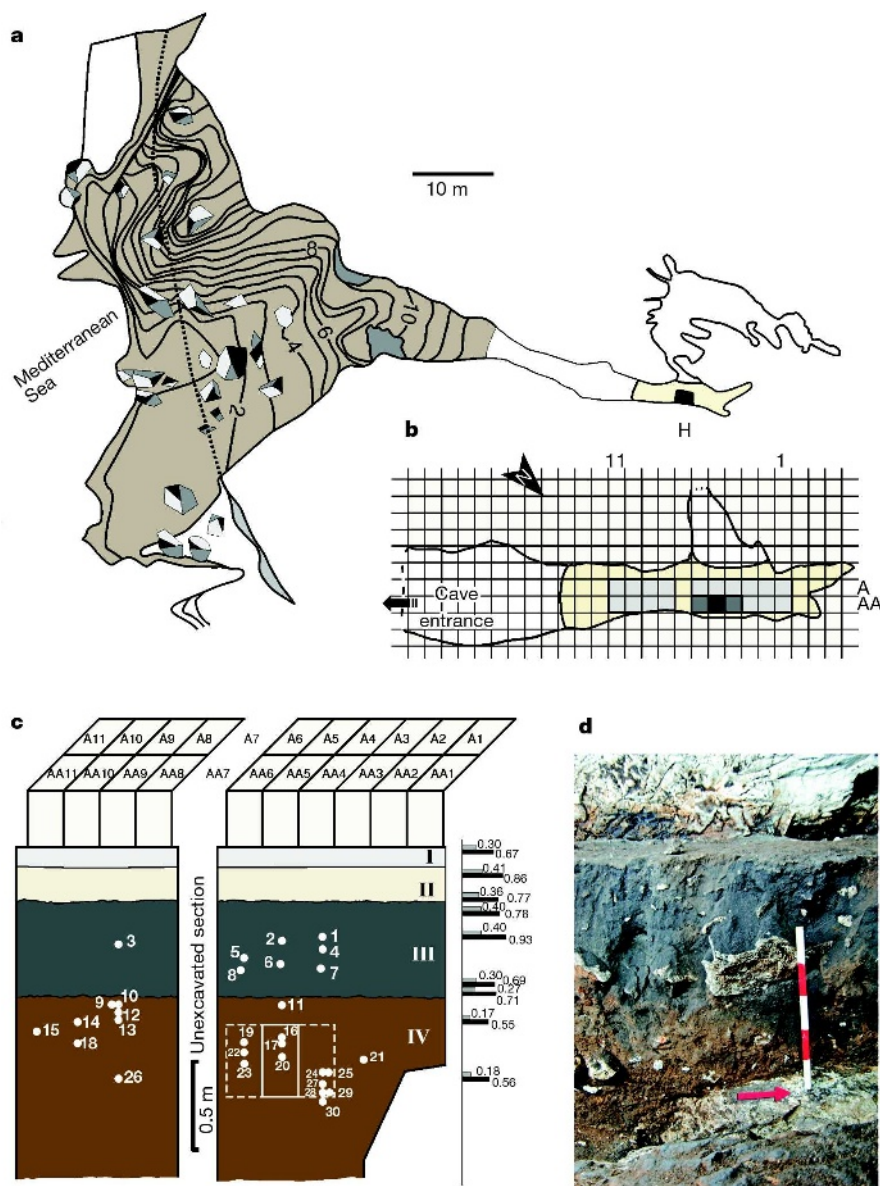


Figure 1 | Plan of Gorham's Cave including the section excavated.

a, General plan of Gorham's Cave showing, at the right, the excavated area in the deep part of the cave (yellow) and the location of the hearth site (black). **b**, Detail of the excavated area (light grey) with the hearth site (black) and the hearth radius of influence (dark grey; see **c**). **c**, Stratigraphy of the excavated inner part of Gorham's Cave described in the text. Numbered dots indicate the positions of charcoal samples; the corresponding AMS dates are given in Table 1. The white rectangle marks the position of the Mousterian hearth site

(see also Supplementary Information). Hatched lines indicate the hearth's radius of influence indicated by charcoal concentrations. The column on the right gives K/Al (black bars) and Mg/Al (grey bars) values for specific points in the stratigraphy (see Supplementary Information for details). The grid above the figure shows the 1×1 m squares excavated. **d**, Photograph showing the section at AA2–AA3. Level IV (brown, below) is clearly differentiated from level III (grey, above). The red arrow indicates north.

Table 1 | The 30 AMS dates in the sequence illustrated in Fig. 1

Sample no.*	Laboratory reference	AMS radiocarbon age (yr)†	$^{13}\text{C}/^{12}\text{C}$ ratio (‰)
1	Beta-181896	13,870 ± 80	-24.0
2	Beta-185343	10,880 ± 80	-25.4
3	Beta-181895	12,460 ± 100	-24.0
4	Beta-184047	12,640 ± 100	-25.4
5	Beta-196780	13,820 ± 100	-24.6
6	Beta-196777	12,540 ± 100	-24.9
7	Beta-181893	16,420 ± 120	-25.5
8	Beta-184042	18,440 ± 160	-21.7
9	Beta-196785	26,070 ± 360	-25.6
10	Beta-196784	28,360 ± 480	-26.1
11	Beta-185344	27,020 ± 480	-25.0
12	Beta-196786	29,910 ± 600	-24.7
13	Beta-196787	31,480 ± 740	-23.7
14	Beta-196792	30,310 ± 620	-24.7
15	Beta-185345	23,780 ± 540	-25.0
16	Beta-196775	24,010 ± 320	-24.0
17	Beta-196773	26,400 ± 440	-23.2
18	Beta-196791	28,570 ± 480	-25.2
19	Beta-196779	29,400 ± 540	-25.4
20	Beta-196776	30,560 ± 720	-24.5
21	Beta-184045	31,110 ± 460	-23.7
22	Beta-196778	29,720 ± 560	-24.8
23	Beta-196782	23,360 ± 320	-22.4
24	Beta-196768	31,290 ± 680	-25.8
25	Beta-196772	31,780 ± 720	-23.1
26	Beta-196789	32,100 ± 800	-24.5
27	Beta-196769	31,850 ± 760	-23.5
28	Beta-196770	28,170 ± 480	-25.9
29	Beta-184048	29,210 ± 380	-25.2
30	Beta-196771	32,560 ± 780	-25.1

Level III dates are numbered 1 to 8; level IV dates are numbered 9 to 30. Three dates from the hearth position are numbered 16, 17 and 20.

* Sample numbers are those used in Fig. 1.

† Errors shown are 2σ .

Carnivore activity is present, but it is not conspicuous (only 5.7% of the remains were affected by carnivore tooth-marks). Post-depositional alterations are not especially destructive; trampling, which produced surface scratches, was the principal factor (7.8%). No bone fragments or artefacts could be refitted. Evidence of re-sedimentation or reworking was not observed either. This confirms that most taxa represented at Gorham's Cave were brought from the surrounding environments by humans and butchered inside the cave with a subsequent negligible access to scavengers.

The sequence of radiocarbon dates presented, including 14 dates at or statistically younger than 30 kyr BP, are the only currently reliable ones that establish the persistence of Neanderthals and associated Mousterian technology after 30 kyr BP. Earlier claims are now dismissed or are uncertain for a variety of reasons and in particular after the revision of dates on bone with the use of ultrafiltration treatment, a treatment only meaningful for dates on bone¹³. Hyaena Den (UK) is now considered older than 30 kyr BP²; the Vindija (Croatia) Neanderthals have been re-dated to between 32 and 33 kyr BP or older¹⁴; Zafarraya (Spain) is now discarded for several reasons¹⁵; the Mezmaiskaya, Russia, Neanderthal is now dated to at least 36 kyr BP¹⁶. The single AMS date on *Cervus* bone for Caldeirão (Portugal)¹⁷ will require revision and is likely, given the result for Hyaena Den of similar age², to be older than 30 kyr BP. Finally, the single ^{14}C date, from *Patella* shells, from Figueira Brava, Portugal, is not statistically younger than 30 kyr BP¹⁷.

The last Neanderthals that occupied Gorham's Cave had access to a diverse community of plants and vertebrates on the sandy plains, open woodland and shrubland, wetlands, cliffs and coastal environments surrounding the site. Such ecological diversity might have facilitated their long survival. The overall pattern of fauna and vegetation in the late Mousterian level IV is consistent with that occurring outside the cave for the greater part of the Late Pleistocene and indicates Mediterranean glacial refugium conditions with a Thermo/Meso-Mediterranean, subhumid climate (Supplementary Information). It consisted of a broad mosaic of mesothermophilous

plant communities, open parkland habitats and seasonal fresh and brackish water sources (Supplementary Information).

The Neanderthals therefore persisted in Mediterranean environments that had acted as glacial refugia for many species throughout the Quaternary period¹⁸. The presence of Aurignacian technology in Bajondillo, Málaga, about 100 km east of Gibraltar at around 32 kyr BP (ref. 19) and of sporadic Gravettian sites after this²⁰ indicates that the late survival of Neanderthals and the arrival of modern humans was a mosaic process in which pioneer groups of moderns and remnant groups of Neanderthals together occupied a highly heterogeneous region for several thousand years. The low density of early Upper Palaeolithic sites in southern Iberia²⁰ and the late presence of Neanderthals, reported in this paper, indicate that for a long time populations of both Neanderthals and modern humans were thinly scattered across the region. The absence of transitional industries that have been attributed to Neanderthal-modern cultural contact²¹, as are found in northern Iberia and France (Châtelperronian), adds weight to the view of limited contact between Neanderthals and moderns in southern Iberia. The transition from the Middle to the Upper Palaeolithic was not, in southern Iberia at least, a sudden rupture but instead took the form of a long and diffuse spatio-temporal mosaic involving populations at low density.

METHODS

Cave deposits often have a complex sedimentation and the possibility of mixing of sediments cannot be discounted. Representative samples were therefore taken from each level at Gorham's Cave and were analysed, both geochemically and mineralogically, with the aim of checking the homogeneity of the stratigraphic levels and the conditions in the cave. The resulting data have been used as proxies to establish a sedimentary regime and conditions within the cavity. The four stratigraphical levels (I–IV) were sampled. Sediment samples were dried and homogenized in an agate mortar for mineralogical and geochemical analyses.

Bulk mineral compositions were obtained by X-ray diffraction (XRD) in accordance with international recommendations²². X-ray diffractograms were obtained with a Philips PW 1710 diffractometer with $\text{CuK}\alpha$ radiation and an automatic slit. Resulting diffractograms were interpreted with Xpovder software²³. Major element measurements (Mg, Al, K, Ca, Mn and Fe) were obtained by atomic absorption spectrometry (AAS; Perkin-Elmer 5100 spectrometer) with an analytic error of 2% in the Analytical Facilities (CIC) of the University of Granada. Al, K, Ca, Ti, Mn, Fe, Cu and Sr have also been quantified with an X-ray fluorescence scanner in the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) laboratories. The XRF-core scanner TATSCAN-F2 (ref. 24) was set to determine bulk intensities of major elements on split sediment sections at intervals of 0.5 cm with an accuracy in standard powder samples of better than 0.20 wt%. A depth correction was performed to compare AAS and XRF-scanner data, indicating a high correlation between techniques. Analyses of trace elements including Ba were performed with inductively coupled plasma-mass spectrometry (ICPMS) after digestion with HNO_3 plus HF. Measurements were performed in triplicate by spectrometry (Perkin-Elmer Sciex Elan 5000), with Re and Rh as internal standards. Variation coefficients determined by the dissolution of ten replicates of powdered samples were more than 3% and 8% for analyte concentrations of 50 and 5 p.p.m., respectively²⁵.

High-resolution geochemical mapping was performed with an XRF scanner in sediments proceeding from level IV (Supplementary Information). Results indicate a sublaminal sedimentation slightly turbated by limestone dropstone, flint tools, fossil bones and carbonate nodules. This confirms that sedimentation in level IV does not correspond to 'breccia' or chaotic deposits.

The 30 AMS dates presented in this paper were analysed by Beta Analytic Inc. All samples were pretreated to eliminate secondary carbon components. The samples were first gently crushed and dispersed in deionized water. They were then given acid washes with hot HCl to eliminate carbonates, and alkali washes with NaOH to remove secondary organic acids. The alkali washes were followed by a final acid rinse to neutralize the solution before drying. Chemical concentrations, temperatures, exposure times and number of repetitions, were applied according to the scarcity of the sample. Each chemical solution was neutralized before application of the next. During these serial rinses, mechanical contaminants such as associated sediments and rootlets were eliminated. This type of pretreatment is considered a 'full pretreatment'.

Received 12 April; accepted 25 August 2006.

Published online 13 September 2006.

1. Brown, P. *et al.* A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
2. Mellars, P. A. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* **439**, 931–935 (2006).
3. Busk, G. On a very ancient cranium from Gibraltar. *Report of the 34th Meeting of the British Association for the Advancement of Science (Bath, 1864)* 91–92 (1865).
4. Garrod, D. A. E., Buxton, L. H. D., Elliot Smith, G. & Bate, D. M. A. Excavation of a Mousterian Rock-shelter at Devil's Tower, Gibraltar. *J. R. Anthropol. Inst.* **58**, 91–113 (1928).
5. Mellars, P. *The Neanderthal Legacy* (Princeton Univ. Press, Princeton, 1996).
6. Waechter, J. D'A. The excavations at Gorham's Cave, Gibraltar, 1951–1954. *Bull. Inst. Archaeol. Lond.* **4**, 189–221 (1964).
7. Barton, N. in *Neanderthals on the Edge* (eds Stringer, C. B., Barton, R. N. E. & Finlayson, C.) 211–220 (Oxbow Books, Oxford, 2000).
8. Pettitt, P. B. & Bailey, R. M. in *Neanderthals on the Edge* (eds Stringer, C. B., Barton, R. N. E. & Finlayson, C.) 155–162 (Oxbow Books, Oxford, 2000).
9. Reimer, P. J. *et al.* Comment on 'Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired 230Th/234U/238U and 14C dates on pristine corals' by R. G. Fairbanks *et al.* (*Quaternary Science Reviews* **24** (2005) 1781–1796) and 'Extending the radiocarbon calibration beyond 26,000 years before present using fossil corals' by T.-C. Chin *et al.* (*Quaternary Science Reviews* **24** (2005) 1797–1808). *Quat. Sci. Rev.* **25**, 855–862 (2006).
10. Romano, F. P. *et al.* Quantitative non-destructive determination of trace elements in archaeological pottery using a portable beam stability-controlled XRF spectrometer. *XRay Spectrom.* **35**, 1–7 (2006).
11. Gaudzinski, S. (2004) A matter of high resolution? The Eemian Interglacial (OIS 5e) in North-central Europe and Middle Palaeolithic subsistence. *Int. J. Osteoarchaeol.* **14**, 201–211 (2004).
12. Grayson, D. K. & Delpéch, F. Ungulates and the Middle-to-Upper Paleolithic transition at Grotte XVI (Dordogne, France). *J. Archaeol. Sci.* **30**, 1633–1648 (2003).
13. Higham, T. F. G., Jacobi, R. M. & Bronk Ramsay, C. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**, 179–195 (2006).
14. Higham, T., Bronk Ramsey, C., Karavanic, I., Smith, F. H. & Trinkaus, E. Revised direct radiocarbon dating of the Vindija G1 Upper Palaeolithic Neanderthals. *Proc. Natl Acad. Sci. USA* **103**, 553–557 (2006).
15. Barroso Ruiz, C. (ed.) *El Pleistoceno Superior de la Cueva del Boquete de Zafarraya* (Junta de Andalucía, Arqueología Monografías, 2003).
16. Skinner, A. R. *et al.* ESR dating at Mezmaiskaya Cave, Russia. *Appl. Radiat. Isot.* **62**, 219–224 (2005).
17. Zilhao, J. in *Neanderthals on the Edge* (eds Stringer, C. B., Barton, R. N. E. & Finlayson, C.) 111–119 (Oxbow Books, Oxford, 2000).
18. Finlayson, C. Biogeography and evolution of the genus *Homo*. *Trends Ecol. Evol.* **20**, 457–463 (2005).
19. d'Errico, F. & Sánchez Goñi, M. F. Neanderthal extinction and the millennial scale climatic variability of OIS 3. *Quat. Sci. Rev.* **22**, 769–788 (2003).
20. Finlayson, C. *Neanderthals and Modern Humans* (Cambridge Univ. Press, Cambridge, 2004).
21. Mellars, P. The Neanderthal problem continued. *Curr. Anthropol.* **40**, 341–364 (1999).
22. Kirsch, H. J. Illite crystallinity: recommendations on sample preparation, X-ray diffraction settings, and interlaboratory samples. *J. Metamorphic Geol.* **9**, 665–670 (1991).
23. Martin, J. D. Qualitative and quantitative powder X-ray diffraction analysis. <<http://www.xpowder.com>> (2004).
24. Sakamoto, T. *et al.* 2006 non-destructive X-ray fluorescence (XRF) core imaging scanner, 'TATSCAN-F2' for the IODP science, scientific drilling. *Integrated ODP* **2**, 37–39 (2006).
25. Bea, F., Montero, P., Stroh, A. & Baasner, J. Microanalysis of minerals by an Excimer UV-LA-ICP-MS system. *Chem. Geol.* **133**, 145–156 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all those who have participated in this project. The project Palaeomed was co-funded by the Government of Gibraltar and the European Union Interreg IIIB Programme Medocc. Palynological and geochemical investigations were funded by Fundación Séneca, Murcia, Spain, and Ministerio de Educación y Ciencia, DGI, Spain, respectively. Geomorphological work was funded by Ministerio de Educación y Ciencia, DGI, Spain.

Author Contributions C.F., F.G.P., J.R.-V. and C.B.S. were responsible for project direction and management. C.F. coordinated the writing. Multidisciplinary work was divided as follows: archaeology and curation—F.G.P., J.M.G.L., A.S.P., J.B.P., F.G.G., K.B., C.A.V. and A.V.; geomorphology and sedimentology—J.R.-V., F.J.J.E., F.M.R. and T.S.; taphonomy—I.C. and Y.F.J.; palaeoecology—C.F., D.A.F. (general), G.F. (palaeolandscapes), E.A. (anthracology), J.S.C., P.L., J.A.L.S., N.F. (palaeobotany), C.P.G.-O. (herpetological taxonomy), J.A.R.C. (mammalian taxonomy) and A.S.M. (avian taxonomy). All authors discussed and interpreted the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.F. (jcfinlay@gibraltar.gi).

LETTERS

Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink

Helge Niemann^{1,2*}, Tina Lösekann^{1*}, Dirk de Beer¹, Marcus Elvert^{1†}, Thierry Nadalig^{3†}, Katrin Knittel¹, Rudolf Amann¹, Eberhard J. Sauter², Michael Schlüter², Michael Klages², Jean Paul Foucher³ & Antje Boetius^{1,2,4}

Mud volcanism is an important natural source of the greenhouse gas methane to the hydrosphere and atmosphere^{1,2}. Recent investigations show that the number of active submarine mud volcanoes might be much higher than anticipated (for example, see refs 3–5), and that gas emitted from deep-sea seeps might reach the upper mixed ocean^{6–8}. Unfortunately, global methane emission from active submarine mud volcanoes cannot be quantified because their number and gas release are unknown⁹. It is also unclear how efficiently methane-oxidizing microorganisms remove methane. Here we investigate the methane-emitting Haakon Mosby Mud Volcano (HMMV, Barents Sea, 72° N, 14° 44' E; 1,250 m water depth) to provide quantitative estimates of the *in situ* composition, distribution and activity of methanotrophs in relation to gas emission. The HMMV hosts three key communities: aerobic methanotrophic bacteria (*Methylococcales*), anaerobic methanotrophic archaea (ANME-2) thriving below siboglinid tubeworms, and a previously undescribed clade of archaea (ANME-3) associated with bacterial mats. We found that the upward flow of sulphate- and oxygen-free mud volcano fluids restricts the availability of these electron acceptors for methane oxidation, and hence the habitat range of methanotrophs. This mechanism limits the capacity of the microbial methane filter at active marine mud volcanoes to <40% of the total flux.

The HMMV (Fig. 1), a circular structure of 1 km diameter and <10 m elevation above the adjacent sea floor, has been studied since the 1990s as a typical example of an active mud volcano⁹. Its formation might have coincided with a submarine landslide during the late Pleistocene, 330,000–200,000 years ago¹⁰. Today, fluids, gas and muds rise from a depth of 2–3 km through a conduit below the HMMV^{10,11}. The emitted gas is of a mixed microbial/thermogenic origin and consists of >99% CH₄ with a $\delta^{13}\text{C}$ -isotope signature of –60‰ (refs 12, 13). The rising fluids are depleted in sulphate, chloride and magnesium as a result of subsurface clay dewatering¹¹. Investigation of the HMMV with RV *Polarstern* and ROV *Victor 6000* in 2003 showed extensive outcroppings of fresh subsurface muds associated with steep thermal gradients¹⁴, gas and fluid vents, and a large gas plume reaching the mixed upper water column above the HMMV^{8,12}. Seafloor videography in combination with geochemical measurements provided *in situ* estimates of gas flux⁸, fluid flow¹⁵ and habitat distribution¹⁶. We focused on the three main concentric habitats above the gassy muds (Fig. 2): the centre of the HMMV, which was devoid of visible epifauna; thiotrophic bacterial mats dominated by a *Beggiatoa* species; and surrounding fields of siboglinid tubeworms. The concentrations of gases in sediments and bottom water were elevated in all three habitats (Table 1). An essential difference between the habitats is the fluid flow modelled from *in situ*

gradients of temperature, oxygen and sulphide concentrations¹⁵. The results indicate that the upward transport of oxidant-depleted mud volcano fluids hinders downward diffusion of oxygen and sulphate, and limits methanotrophic habitats in the centre and below bacterial mats to the uppermost millimetres to centimetres, respectively¹⁵. We have now identified and quantified the methanotrophs that populate the three habitats to assess their role in controlling methane flux at the HMMV (Table 1; Figs 2, 3).

The surface of the most recent mud flows in the centre of the HMMV, characterized by steep temperature gradients of 3°C m^{–1} (ref. 14), hosted high numbers of microbes, reaching 3.6 (±2.1; ±s.d.) × 10⁹ cells cm^{–3}, of which 56% (±8%) belonged to type I aerobic methanotrophic bacteria of the orders *Methylococcales* and *Methylophaga* (Figs 2b, 3c). The 16S ribosomal RNA gene sequences of the dominant *Methylococcales* types (HMMV-MetI and -MetII) are closely related to sequences of methanotrophic mussel symbionts (94–97%)¹⁷ and to clone sequence Hyd24-1 (96–99%) from gas-hydrate-bearing sediments of Hydrate Ridge¹⁸. High concentrations of a ¹³C-depleted (–80‰), type I methanotroph-specific fatty acid (C16:1ω8c) were found in the same samples (Fig. 3d). As predicted by the fluid flow model for this area¹⁵, methanotroph cell numbers, lipid biomarkers and rates of aerobic methane oxidation (MOx) peaked in the uppermost surface sediment and decreased more than fivefold in the second centimetre because of limited oxygen penetration (Fig. 3a–d). Only small amounts of methanotroph lipids (<0.1 µg per gram sediment dry weight; µg g-dw^{–1}) and very low numbers of cells (~10⁷ cells cm^{–3}) were found below a sediment depth of 5 cm (Fig. 3c, d). ANME cells were not microscopically detectable in the centre cores using all known ANME oligonucleotide probes, and anaerobic oxidation of methane (AOM) was not detectable.

In situ microsensor profiles of sulphide concentrations in HMMV sediments indicated that AOM occurs in older mud flows, 2–3 cm below the sulphide-oxidizer mats (Figs 2c, 3f). Accordingly, microbial biomass peaked in the top 3 cm (Fig. 3g) and mainly consisted of a previously undescribed consortium of archaea and sulphate-reducing bacteria (SRB) forming dense cell aggregates (Fig. 2c). 16S rRNA gene analyses of DNA revealed a dominant cluster of archaeal sequences forming a new phylogenetic group named ANME-3 (Supplementary Fig. 1). Single sequences of this cluster were only sporadically detected at cold seeps^{19,20}, which are typically dominated by other phylogenetic clades of anaerobic methanotrophs (that is, ANME-1 and ANME-2 with their bacterial partners, sulphate reducers of the *Desulfosarcina/Desulfococcus* (DSS) branch)^{19–21}. We propose that members of the ANME-3 cluster are also capable of AOM, for three reasons: the AOM maximum coincides with the sulphide

¹Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany. ²Alfred Wegener Institute for Polar and Marine Research, 27515 Bremerhaven, Germany. ³Centre Ifremer de Brest, BP70, 29280 Plouzané, France. ⁴International University Bremen, 28759 Bremen, Germany. [†]Present addresses: DFG Research Center Ocean Margins, University of Bremen, 28334 Bremen, Germany (M.E.); UMR 7156 Université Louis-Pasteur/CNRS, Département Microorganismes, Génomes, Environnement, 67083 Strasbourg Cedex, France (T.N.).

*These authors contributed equally to this work.

production zone; there are high numbers of ANME-3 aggregates ($19.7 (\pm 0.6) \times 10^6$ aggregates cm^{-3} at 1–2 cm sediment depth), making up 80% of total cell numbers; and concentrations of the lipid biomarkers *sn*2-hydroxyarchaeol, archaeol and specific penta-methylcosenes are highly elevated (Fig. 3e–h). Moreover, the archaeal lipid biomarkers have stable C-isotope signatures ($\delta^{13}\text{C}$ -values $< -98\text{‰}$), indicating that methane is the main carbon source of ANME-3 cells. Analyses of 16S rRNA genes and lipid biomarkers showed that the bacterial partner associated with ANME-3 is different from all other ANME consortia (see Supplementary Information). It is closely related to *Desulfobulbus* spp. (DBB) and also takes up methane-derived carbon, as indicated by its specific lipid biomarker C17:1 ω 6c with a $\delta^{13}\text{C}$ -value of -70‰ (Fig. 3h). As predicted by the fluid flow model, 95% of ANME-3/DBB aggregates were found directly below the *Beggiatoa* mats (Fig. 3g). Because of reduced fluid flow velocities, sulphate can penetrate the upper 4–6 cm of the sediment column (Fig. 3f, Table 1). Furthermore, the re-oxidation of

sulphide by the sulphide oxidizer mats replenishes the sulphate pool in these sediments, further contributing to the maintenance of high activity and biomass of the AOM community in this zone.

The next significant transition in HMMV habitats occurs about 300–400 m from its geographical centre, where the sulphide oxidizer mats are replaced by dense tubeworm colonies on the hummocky periphery of the HMMV (Fig. 2d, Table 1). Two siboglinid tubeworm species, *Oligobranchia haakonmosbiensis* and *Sclerolium contortum*¹³, populate this area with biomasses of 1–2 kg wet weight m^{-2} (estimated from sieved boxcore samples). *Oligobranchia haakonmosbiensis*, the dominant species in the investigated area, is up to 60 cm long and has a diameter of 0.5 mm. In the tubeworm fields, fluid flow rates modelled from temperature profiles (0.4 m yr^{-1}) were lower than those in the centre or bacterial mat area¹⁴. In agreement with the low subsurface temperature gradient in this zone¹⁴ (Fig. 3e, i), gravity cores from this area contained gas hydrate below the worm-infested sediment layers. Balancing downward diffusion and upward fluid flow, maximum sulphate penetration could reach 12 cm. However, the measured oxygen and sulphate penetration in the tubeworm field (Figs 2d, 3i, j; Table 1) show that the tubeworms substantially increase the transport of electron acceptors into deeper sediment layers. Accordingly, the maximum AOM was found between the base of the worm tubes and the hydrate layer (60–90 cm below sea floor (b.s.f.)), coinciding with a subsurface peak of ANME-2/DSS aggregates (5.5×10^6 aggregates cm^{-3}) and their specific biomarker lipids (Fig. 3i–l). Integrated *ex situ* methane consumption was higher than in other zones of the HMMV (Table 1). Microbe–invertebrate symbioses have an advantage over free-living microbial populations because they can engineer their environment to increase access to both electron donors and acceptors by special migratory behaviour, mining and pumping. An interesting question remains: which factors determine the transition between *Beggiatoa* mats and tubeworm fields, and the temporal succession of the thio- and methanotrophic communities? Visual observation of the sea floor showed that sulphide-oxidizing mats and tubeworm colonies are mutually exclusive (no worms were found within bacterial mats). Tubeworm larvae probably cannot settle on *Beggiatoa* mats owing to the high fluid flow and high AOM-based sulphide concentration in this habitat. However, compared with filamentous sulphide oxidizers, tubeworms can reach much deeper zones of sulphide production, and might efficiently compete for sulphide and methane uptake.

In conclusion, we found that methane oxidation was the main energy source for microbial biomass in all habitats of the HMMV. Populations of aerobic or anaerobic methanotrophs comprised 56–76% of the total microbial community (centre and *Beggiatoa* mat habitats). At their maxima in abundance and activity (Fig. 3), methanotrophic communities show relatively low²² cell-specific rates of methane oxidation (around 0.1 fmol CH_4 per cell per day, under atmospheric pressure), possibly owing to the low ambient temperature of -1°C . The total integrated activity and biomass was lower for the aerobic methanotrophs than for the AOM communities, as the former were limited to the surface of the centre muds by high fluid flow (Table 1, and see Supplementary Information). ANME-3 and ANME-2 consortia both dominate areas with lower flow, and profit from the association with other organisms that re-oxidize sulphide to sulphate. Total methane consumption in the centre, *Beggiatoa* and tubeworm habitats was up to 0.2, 1.8 and $3.0 \times 10^6 \text{ mol yr}^{-1}$, respectively (sum $5.0 \times 10^6 \text{ mol yr}^{-1}$). In 2003, methane emission to the hydrosphere from the HMMV reached $8\text{--}35 \times 10^6 \text{ mol yr}^{-1}$ (ref. 8), resulting in a total methane flux of $13\text{--}40 \times 10^6 \text{ mol yr}^{-1}$. Aerobic oxidation of methane had a minor role (1–3%) compared with anaerobic oxidation, which removed up to 37% of the total methane flux. Hence, the efficiency of microbial methane removal is substantially lower than in other methane-fuelled systems^{23–25}. These data confirm that fluid advection acts as a main habitat-structuring factor at cold seep ecosystems, by

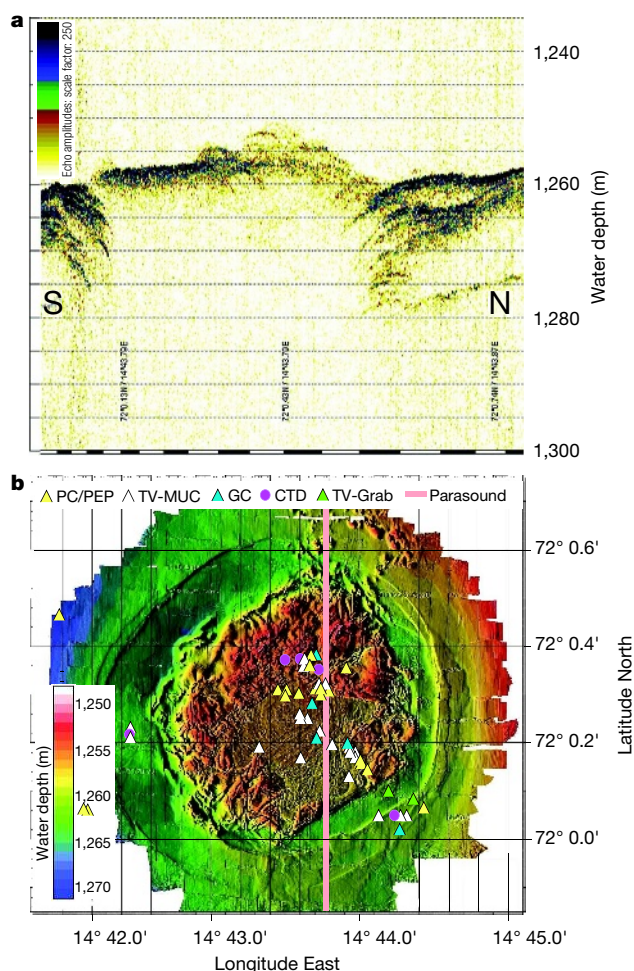


Figure 1 | The Haakon Mosby Mud Volcano. **a**, Sediment echosounder image (PARASOUND) of a 1.8-km-long profile crossing the mud volcano from south to north (pink line in **b**, $14^\circ 43.8' \text{ E}$). High gas content in the centre sediments considerably decreases signal penetration. The stacks of sedimentary layers of the northern and southern rims might correspond to former mudflows. Black and white bars along the bottom represent 100-m horizontal distances. The y-axis (height) is exaggerated 24-fold. **b**, High-resolution, bathymetric map of the HMMV showing sampling stations for ROV push cores (PC), TV-guided multiple corer cores (TV-MUC), gravity corer (GC), TV-guided grab (TV-Grab), ROV PEP bottles (PEP) and CTD rosette (CTD). The IFREMER software CARAIBES was used to process the microbathymetry data acquired by the multibeam echosounder (Reson SeaBat 8125 and EM2000, georeferenced by Posidonia USBL positioning) mounted on ROV Victor 6000.

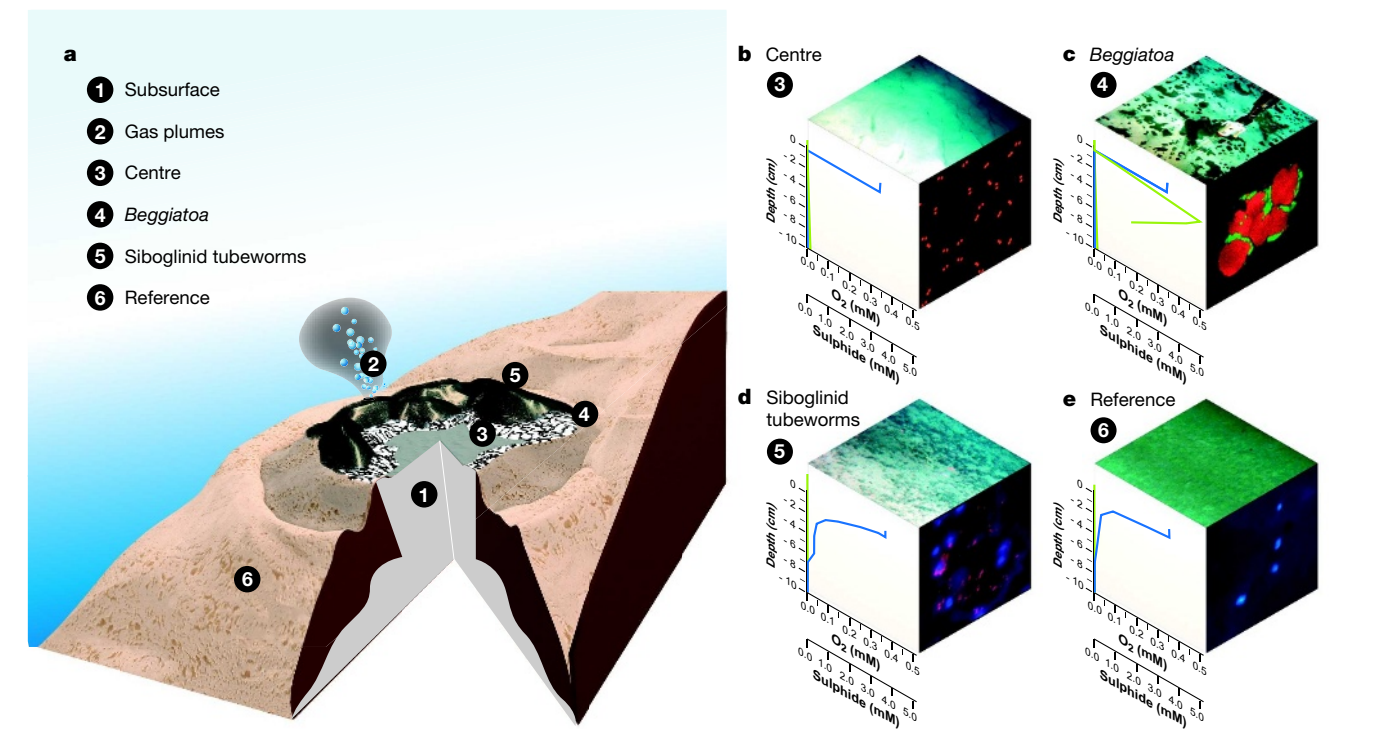


Figure 2 | A schematic diagram of the different microbial habitats at the HMMV. **a**, Overall schematic. Vertical height is exaggerated 50-fold. **b–e**, Detail of specific habitats. Images of the sea floor (top panel), idealized concentration gradients (left panel) of oxygen (blue) and sulphide (green), and micrographs of the dominating microbial organisms (right panel; FISH probe specifications are provided in Supplementary Information) are shown as three-dimensional structures. **a**, (1) Subsurface muds rising from a deep

reservoir. (2) Gas and fluid release in the northern centre. **a**, **b**, (3) Central sediments colonized by aerobic methanotrophic bacteria (type I methanotrophs). **a**, **c**, (4) *Beggiatoa* mats covering sediments dominated by an ANME-3 (red)/DBB (green) population. **a**, **d**, (5) Colonies of siboglinid tubeworms harbouring symbiotic bacteria (shown in pink; host nuclei shown in blue). **a**, **e**, (6) Reference station (Micrograph: DAPI staining).

regulating the availability of electron acceptors and hence the distribution and activity of methanotrophs.

It is generally assumed that with increasing upward flow velocities, more dissolved methane is transported to the surface, fuelling increasingly active seep communities (for example, see ref. 26).

However, geofluids produced from compaction processes and dewatering are depleted in electron acceptors. High flows of methane-laden fluids introduce a negative switch in the relationship between fluid flow and methanotroph activity. Furthermore, if such fluids transport heat from below, they could even suppress the

Table 1 | Gas consumption and emission at HMMV

	Centre	Sulphide oxidizer mats	Siboglinid tubeworm fields	Reference
Area (km ²) (total = 0.74 km ² ; ref. 16)	0.11	0.22	0.41	-
Methane concentration (μM) (20 cm above sea floor, n = 3)	5.7 (±1.0)	0.3 (±0.1)	0.7 (±0.4)	0.02 (±0.01)
Fluid flow (m yr ⁻¹) (based on temperature gradients) ^{14,15}	1.3–5.3	0.6–1	0.4	NA
Fluid flow (m yr ⁻¹) (based on chemical gradients) ¹⁵	3–6	0.3–0.6	NA	0
Penetration depth (cm) O ₂ (in situ) ¹⁵	0.1–0.3	0.1–0.2	3–10	5
Penetration depth (cm) sulphate (ex situ) ¹⁵	2	4–6	100	>15
Methanotrophs	<i>Methylococcales</i> , <i>Methylophaga</i>	ANME-3/DBB	ANME-2/DSS, ANME-3/DBB	NA
Methanotrophic biomass (mol C m ⁻²)	0.04	0.6	0.4 (80; tubeworm biomass)	NA
Ex situ MOx (mol m ⁻² yr ⁻¹)	0.9 (±0.4)	NA	0.2 (±0.03)*	0.02 (±0.01)
Ex situ AOM (mol m ⁻² yr ⁻¹)	0	4.5 (±1.5)	7.1 (±2.0)*	0.1 (±0.05)
In situ O ₂ consumption (mol m ⁻² yr ⁻¹) ¹⁵	3.8*	4.1	0.6 (without tubeworm respiration)	0.3
In situ sulphide production (mol m ⁻² yr ⁻¹) ¹⁵	0	8.2*	NA	0
Areal methane consumption (×106 mol yr ⁻¹)	0.2	1.8	3.0	-
Relative methane consumption (% of total flux, 13–40 × 106 mol yr ⁻¹)	1–2	5–14	7–24	-

Total gas emission in 2003 was 8–35 × 10⁶ mol CH₄ yr⁻¹ based on *in situ* measurements at three active vents of the HMMV⁸. Maximal relative methane consumption by microbial communities was estimated from microbial methane consumption* divided by the total flux. All values are ±s.e.m. NA, not available.

*Values used for calculations of real and relative methane consumption.

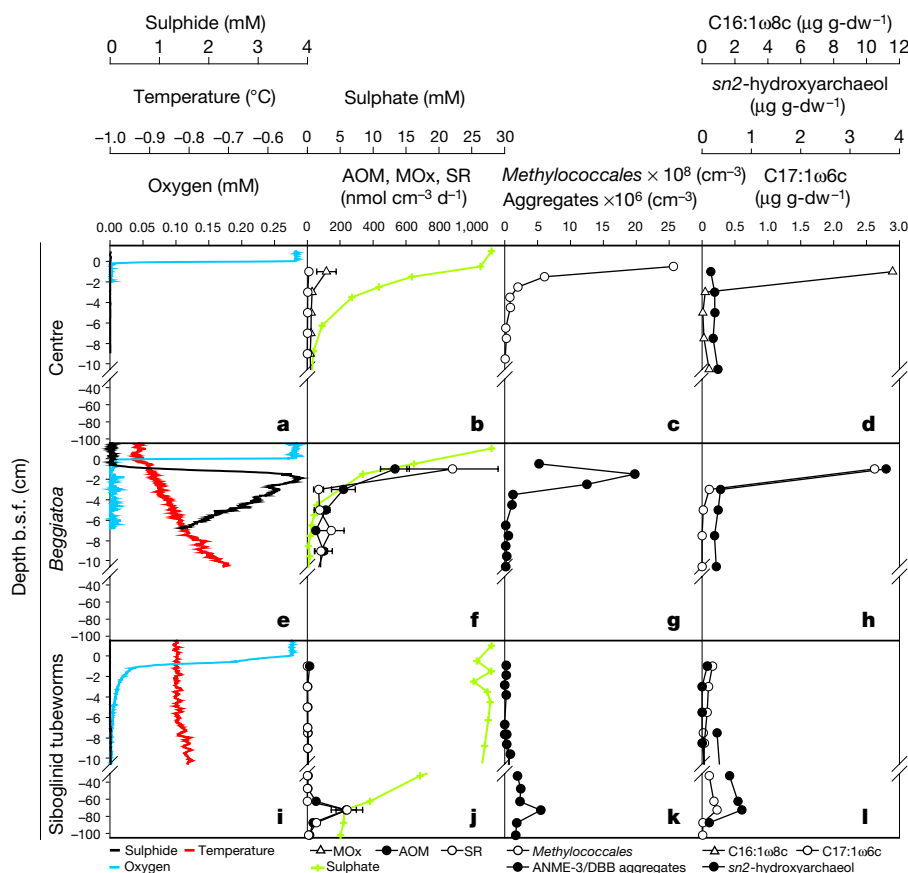


Figure 3 | Vertical distribution of methanotrophs in relation to environmental factors. Profiles of oxygen, sulphide and temperature (a, e, i); sulphate, *ex situ* methane and sulphate turnover (b, f, j); bacteria and aggregate counts (c, g, k), and selected lipid biomarkers (d, h, l) from the centre (a–d), *Beggiatoa*-covered sediments (e–h), and the tubeworm field

(i–l). Oxygen, sulphide and temperature were measured *in situ* with microsensors mounted on a deep-sea profiling unit. AOM, SR, sulphate, cell and aggregate counts, and lipid biomarker contents, were determined using ROV pushcorer samples from the same sites. The errors of MOx and AOM measurements (Fig. 3b, f, j) are given as s.e.m.

formation of hydrates, which represent an important—although dynamic—sink for methane on earth²⁷. Consequently, as in the case of the HMMV, methane rising with warm, oxidant-depleted fluids might accumulate in muds and escape as free gas to the hydrosphere. Here we have shown that the efficiency of the microbial filter against methane decreases considerably at fluid flow rates $>0.4 \text{ m yr}^{-1}$, causing increased efflux of methane to the hydrosphere and probably even to the atmosphere⁸. Velocities of decimetres to metres per year are within the average range of fluid flow measured at cold seeps^{28,29}, so the described inhibition of microbial methane consumption might be widespread at seeps. Further quantitative *in situ* measurements of fluid flow and microbial methane consumption are necessary to define the relevance and control of methane emissions from submarine mud volcanoes and other fluid-flow driven geo-bio-systems.

METHODS

Sampling. Our studies took place during two cruises with RV *L'Atalante* (2001) and RV *Polarstern* (2003), both equipped with ROV *VICTOR 6000*. Sediments were collected with ROV pushcores and gravity cores for on board (*ex situ*) measurements. Sediment cores as well as *in situ* microsensor profiles were obtained along a radial transect from the centre to the east of the mud volcano (Fig. 1b). Bottom water was sampled with a bottom water sampler and methane concentrations were determined by gas chromatography⁸. A detailed description of the methods described here is provided in the Supplementary Information.

Sulphate reduction and methane oxidation rates. Microbial rates of aerobic and anaerobic methane oxidation (MOx and AOM, respectively) and sulphate reduction (SR) in sediments were determined *ex situ* using $^{14}\text{CH}_4$ and $^{35}\text{SO}_4^{2-}$ tracers^{22,25}.

Lipid analysis. Lipid biomarkers from frozen sediment samples and tubeworm tissue were extracted, separated and derivatized as described previously^{22,30}.

Single lipid compounds were analysed by gas chromatography, mass spectrometry and isotope ratio mass spectrometry to determine their quantity, identity and stable carbon isotope ratio, respectively³⁰.

Fluorescence *in situ* hybridization (FISH). Cells of *Methylococcales* as well as ANME-3/DBB and ANME-2/DSS aggregates were quantified by FISH with monolabelled oligonucleotide probes as previously described²⁰.

Sulphate and chloride concentrations. Sulphate concentrations were determined by high-performance liquid chromatography (HPLC) analysis from the supernatant of 5 ml sediment fixed with 25 ml zinc acetate solution²⁵. Chloride concentrations were determined from pressure-filtered pore water by ion chromatography.

Microsensor measurements. Microsensors (20 μm tip diameter) for O_2 , H_2S and pH were manufactured and used as described previously¹⁵. Profiles were recorded *in situ* with a spatial resolution of 0.025 cm by deploying a profiler unit (equipped with the sensors) with the ROV or a free falling lander system. A detailed description of *in situ* data and the geochemical models applied here is provided elsewhere¹⁵.

Received 16 July; accepted 31 August 2006.

- Kopf, A. J. Significance of mud volcanism. *Rev. Geophys. B* **40**, 1–49 (2002).
- Milkov, A. V. *et al.* Global gas flux from mud volcanoes: A significant source of fossil methane in the atmosphere and the ocean. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016358 (2003).
- Bohrmann, G. *et al.* Mud volcanoes and gas hydrates in the Black Sea: new data from Dvurechenskii and Odessa mud volcanoes. *Geo-Mar. Lett.* **23**, 239–249 (2003).
- Pinheiro, L. M. *et al.* Mud volcanism in the Gulf of Cadiz: results from the TTR-10 cruise. *Mar. Geol.* **195**, 131–151 (2003).
- Loncke, L. *et al.* Mud volcanoes, gas chimneys, pockmarks and mounds in the Nile deep-sea fan (Eastern Mediterranean): geophysical evidences. *Marine Petrol. Geol.* **21**, 669–689 (2004).
- MacDonald, I. R. *et al.* Transfer of hydrocarbons from natural seeps to the water column and atmosphere. *Geofluids* **2**, 95–107 (2002).

7. Rehder, G. *et al.* Enhanced lifetime of methane bubble streams within the deep ocean. *Geophys. Res. Lett.* **29**, 21–24 (2002).
8. Sauter, E. J. *et al.* Methane discharge from a deep-sea submarine mud volcano into the upper water column by gas hydrate-coated methane bubbles. *Earth Planet. Sci. Lett.* **243**, 354–365 (2006).
9. Vogt, P. R. *et al.* Haakon Mosby mud volcano: A warm methane seep with seafloor hydrates and chemosynthesis-based Ecosystem in late Quaternary Slide Valley, Bear Island Fan, Barents Sea passive margin. *EOS Trans. AGU* **78**, 187–189 (1997).
10. Vogt, P. R., Gardner, J. & Crane, K. The Norwegian-Barents-Svalbard (NBS) continental margin: Introducing a natural laboratory of mass wasting, hydrates, and ascent of sediment, pore water, and methane. *Geo-Mar. Lett.* **19**, 2–21 (1999).
11. Ginsburg, G. D. *et al.* Gas hydrate accumulation at the Håkon Mosby Mud Volcano. *Geo-Mar. Lett.* **19**, 57–67 (1999).
12. Damm, E. & Budéus, G. Fate of vent-derived methane in seawater above the Håkon Mosby mud volcano (Norwegian Sea). *Mar. Chem.* **82**, 1–11 (2003).
13. Milkov, A. V. *et al.* Geological, geochemical, and microbial processes at the hydrate-bearing Håkon Mosby mud volcano: a review. *Chem. Geol.* **205**, 347–366 (2004).
14. Kaul, N., Foucher, J. P. & Heesemann, M. Estimating mud expulsion rates from temperature measurements on Håkon Mosby Mud Volcano, SW Barents Sea. *Mar. Geol.* **229**, 1–14 (2006).
15. de Beer, D. *et al.* In situ fluxes and zonation of microbial activity in surface sediments of the Håkon Mosby Mud Volcano. *Limnol. Oceanogr.* **51**, 1315–1331 (2006).
16. Jerosch, K. *et al.* Spatial distribution of mud flows, chemoautotrophic communities, and biogeochemical habitats at Håkon Mosby Mud Volcano. *Mar. Geol.* (submitted).
17. Distel, D. L., Lee, H. K. W. & Cavanaugh, C. M. Intracellular coexistence of methanotrophic and thioautotrophic bacteria in a hydrothermal vent mussel. *Proc. Natl Acad. Sci. USA* **92**, 9598–9602 (1995).
18. Knittel, K. *et al.* Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia margin, Oregon). *Geomicrobiol. J.* **20**, 269–294 (2003).
19. Orphan, V. J. *et al.* Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl. Environ. Microbiol.* **67**, 1922–1934 (2001).
20. Knittel, K. *et al.* Diversity and distribution of methanotrophic archaea at cold seeps. *Appl. Environ. Microbiol.* **71**, 467–479 (2005).
21. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
22. Niemann, H. *et al.* Methane emission and consumption at a North Sea gas seep (Tommeliten area). *Biogeosciences* **2**, 335–351 (2005).
23. Hinrichs, K.-U. & Boetius, A. in *Ocean Margin Systems* (eds Wefer, G., Billelt, D. & Hebbeln, D.) 457–477 (Springer, Berlin, 2002).
24. Reeburgh, W. S. in *Treatise on Geochemistry* Vol. 4 (ed. Keeling, R. F.) 65–89 (Elsevier–Pergamon, Oxford, 2003).
25. Treude, T. *et al.* Anaerobic oxidation of methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. *Mar. Ecol. Prog. Ser.* **264**, 1–14 (2003).
26. Sibuet, M. & Olu, K. Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. *Deep-Sea Res. II* **45**, 517–567 (1998).
27. Judd, A. G. *et al.* The geological methane budget at Continental Margins and its influence on climate change. *Geofluids* **2**, 109–126 (2002).
28. Luff, R. & Wallmann, K. Fluid flow, methane fluxes, carbonate precipitation and biogeochemical turnover in gas hydrate-bearing sediments at Hydrate Ridge, Cascadia Margin: Numerical modeling and mass balances. *Geochim. Cosmochim. Acta* **67**, 3403–3421 (2003).
29. Brown, K. M. *et al.* Correlated transient fluid pulsing and seismic tremor in the Costa Rica subduction zone. *Earth Planet. Sci. Lett.* **238**, 189–203 (2005).
30. Elvert, M. *et al.* Characterization of specific membrane fatty acids as chemotaxonomic markers for sulfate-reducing bacteria involved in anaerobic oxidation of methane. *Geomicrobiol. J.* **20**, 403–419 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The expeditions 'AWI' on RV *L'Atalante* in 2001 and 'ARK XIX3b' on RV *Polarstern* in 2003, both with ROV *Victor 6000*, were jointly planned, coordinated and carried out by the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany, and the French Research Institute for Exploitation of the Sea (IFREMER) in Brest, France. We thank the captain and crew, the team of the ROV *Victor 6000* and the shipboard scientific community of the RV *Polarstern* and *L'Atalante* for their help at sea. We thank S. Lüdeling, L. Baumann, V. Beier, I. Busse, F. Heinrich, G. Eickert, A. Nordhausen, J. Rogenhagen and R. Usbeck for technical assistance, C. Edy for help with georeferencing, U. Witte for providing the benthic lander technology, and S. Joye for comments on the manuscript. This is publication no. GEOTECH-235 of the R&D-Programme GEOTECHNOLOGIEN, Project MUMM, funded by the German Ministry of Education and Research (BMBF) and German Research Foundation (DFG).

Author Contributions T.L., T.N., K.K. and R.A. carried out the 16S-rRNA-based analyses and microscopy, H.N. and M.E. the lipid biomarker work, A.B. and H.N. the rate measurements, D.B., E.S. and M.S. the microsensor and geochemical measurements, and M.K., M.S., A.B., H.N. and J.P.F. the geo- and video-graphical survey and experimental strategy. A.B., H.N. and D.B. wrote the manuscript text, and H.N., T.L. and K.K. wrote the supplements. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequence data have been deposited in EMBL, GenBank and the DDBJ nucleotide sequence database under accession numbers AJ704650–AJ704653, AJ704631 and AM287206–AM287207. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (aboetius@mpi-bremen.de).

Fishing elevates variability in the abundance of exploited species

Chih-hao Hsieh¹, Christian S. Reiss², John R. Hunter¹, John R. Beddington³, Robert M. May⁴ & George Sugihara¹

The separation of the effects of environmental variability from the impacts of fishing has been elusive, but is essential for sound fisheries management^{1–7}. We distinguish environmental effects from fishing effects by comparing the temporal variability of exploited versus unexploited fish stocks living in the same environments. Using the unique suite of 50-year-long larval fish surveys from the California Cooperative Oceanic Fisheries Investigations⁴ we analyse fishing as a treatment effect in a long-term ecological experiment. Here we present evidence from the marine environment that exploited species exhibit higher temporal variability in abundance than unexploited species. This remains true after accounting for life-history effects, abundance, ecological traits and phylogeny. The increased variability of exploited populations is probably caused by fishery-induced truncation of the age structure, which reduces the capacity of populations to buffer environmental events^{1,5,8,9}. Therefore, to avoid collapse, fisheries must be managed not only to sustain the total viable biomass but also to prevent the significant truncation of age structure^{1,5,8,9}. The double jeopardy of fishing to potentially deplete stock sizes and, more immediately, to amplify the peaks and valleys of population variability⁷, calls for a precautionary management approach^{10,11}.

One of the most difficult issues in the management of fisheries is to evaluate the effects of fishing in the context of a changing environment^{2,3,6}. Statistics from the Food and Agriculture Organization of the United Nations¹² and recent studies^{3,6,13,14} indicate that many commercially important fish populations have been declining in the past several decades. However, the extent to which such declines are due to fishing, to environmental change, or to some combination of these effects remains a matter of debate^{15–17}.

Understanding how fishing and environmental variability interact to produce an effect on exploited populations (commercially targeted species) is an unsolved problem in fisheries science^{7,18}. For example, it is not known how fishing and environmental variability can either magnify or diminish management risk by affecting the variability (resilience) of exploited populations^{7,18}. Lack of understanding of the sources of temporal variability in fish abundance affects biological reference points, decision making and risk assessment in precautionary fisheries management¹⁹, and determination of extinction risk for populations²⁰; it is largely responsible for the uncertainty and indecision that have led to past failures in managing fisheries²¹. In many historical cases, fisheries management has failed to recognize possible fishing effects until population collapse^{2,5,21}.

We show that fishing effects may appear even in the absence of significant declining trends in populations. We find that fishing increases population variability (fluctuations of populations through time) and that this is an indication that fishing is having a negative impact on populations that is not yet reflected in declining abundance. We evaluate fishing effects by comparing temporal variability in

the larval abundance of exploited species to that of unexploited species living in the same ecosystem. Larval abundance is an established proxy for adult biomass (see Methods and Supplementary Data). Although no environmental variables are directly examined, environmental effects are implicit in our measurements of the temporal variability of fish populations¹¹.

Whether fishing will increase or decrease the population variability of exploited species is a long-standing debate^{7,18}. The discussion (the origins of which trace back at least to the 1970s) remained theoretical because no data existed to resolve the controversy. These data are now available from the California Cooperative Oceanic Fisheries Investigations (CalCOFI) larval fish time series (1951–2002) for the California Current System. We use the coefficient of variation of annual larval abundance to represent temporal variability for the adults of 29 coastal or neritic species (13 exploited and 16 unexploited species) that were abundant and consistently enumerated within CalCOFI⁴ (Supplementary Data Table S2).

Fishing is a selective process, thus the exploited and unexploited groups may not be random with respect to other potential explanatory variables that could relate to the coefficient of variation. In addition to fishing there could be systematic differences with respect to other factors such as life-history traits that could influence the coefficient of variation. Therefore, to isolate the effect of fishing as an explanatory variable, we use multiple regression analysis to factor out variables that could produce biases associated with fishing. These variables include life-history effects, abundance, ecological traits and phylogeny (see Methods). We consider life-history traits that are known to influence population responses to fishing and the environment²². In theory^{20,22–24}, the coefficient of variation of annual larval abundance is negatively related to maximum length, length at maturation, age at maturation, spawning duration and trophic level, and positively related to fecundity. Abundance is included in the regression model because higher variability could be statistically associated with higher abundance²⁰. Finally, ecological traits (geographic region, habitat and spawning mode) and phylogenetic constraints are examined (see Methods and Supplementary Data).

When data from exploited and unexploited species are grouped together, two of the six life-history traits fail to show the expected relationships with the coefficients of variation of annual larval abundance (Fig. 2 of Supplementary Data). The coefficients of variation are positively correlated with maximum length and length at maturation whereas theory predicts negative correlations^{20,22–24}. However, when the analyses are performed separately on exploited and unexploited groups, the relationships between the coefficients of variation and life-history traits follow the theoretical predictions (Fig. 3 of Supplementary Data). Except for fecundity (for which data for unexploited species are too sparse to be conclusive, $n = 4$), the predicted relationship between life-history traits and coefficients of

¹Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0202, USA. ²Southwest Fisheries Science Center, National Marine Fisheries Service, 8604 La Jolla Shores Drive, La Jolla, California 92037-1508, USA. ³Division of Biology, Faculty of Natural Sciences, Imperial College London, RSM Building, South Kensington campus, London SW7 2AZ, UK. ⁴Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

variation is stronger for the unexploited species data set. The relationship between life-history traits and coefficients of variation is weaker for the exploited species data set, perhaps because fishing pressure varies among species and this in itself might obscure the predicted relationships (see Supplementary Data).

The most important result is that, after accounting for life-history effects, abundance, ecological traits and phylogeny, the coefficients of variation of annual larval abundance for exploited species are significantly higher than those for unexploited species. When accounting for all effects, the overall significance of the multiple regression is $P < 0.001$. Notably, age at maturation (Fig. 1) and geographic region (Fig. 2a) emerge as the only two significant factors in the full model ($P = 0.015$ and $P < 0.001$, respectively). Age at maturation is probably the best proxy for generation time, which should be important in determining population variability. The remaining life-history traits and abundance are highly correlated with age at maturation, so it is not surprising to find them redundant (and eliminated) in the regression model. Warm-water species exhibit higher coefficients of variation than cool-water and widely distributed species (Fig. 2). Phylogeny is not a significant variable affecting the coefficients of variation (see Fig. 5 of Supplementary Data).

Could the higher variability in larval abundance of the exploited species be caused by a long-term declining trend in their abundance due to exploitation? Contrary to this speculation, we found no systematic differences between the exploited and unexploited groups in the prevalence of declining trends in the 50-yr interval (Table 1). Out of 13 exploited species, only 2 had significant declines, which is similar to the fraction (2 out of 16) of unexploited species that showed significant declines. However, for safe measure, we recalculated coefficients of variation after removing low-frequency trends in the abundance time series, and found that the exploited species still exhibit higher variability after accounting for our ancillary variables (see Table 5 of Supplementary Data). All these results indicate that the exploited (fished) species are more variable in terms of abundance than the unexploited species.

How might fisheries cause an increase in variability in the abundance of exploited populations? As suggested nearly 30 yr ago⁷, in many fish populations the main source of variability lies in recruitment: the transition from the larval stage to the adult stage. This can be shown in a simple population model:

$$N_{t+1} = N_t e^{(-M-F)} + R_t$$

where N_t is the adult abundance and R_t is the recruitment at time t , and M and F are natural and fishing mortality, respectively. Clearly,

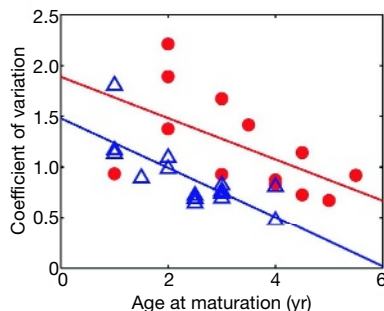


Figure 1 | Relationships between coefficients of variation of annual larval abundance and age at maturation for exploited and unexploited species.

The correlations between coefficients of variation and age at maturation were significant for both exploited (red, filled circles) and unexploited (blue, open triangles) groups ($P < 0.05$). The coefficients of variation for the exploited group were higher than those for the unexploited group, even after accounting for the effect of age at maturation ($P = 0.002$). For graphical presentation, missing data on age at maturation for five species were imputed using the expectation maximization algorithm³⁰ (for species names, see Fig. 4 of Supplementary Data).

as the level of fishing mortality increases, the population dynamics are increasingly dominated by recruitment, and at the limit the population variability is equal to the variability of recruitment. Thus, one would expect fishing to increase population variability.

However, beyond this effect, recruitment variability itself should be further amplified through the effect that fisheries have on truncating the size–age structure of a population^{1,5,8,9}. It is believed that fisheries operate by selectively removing large and old individuals through size–age selective fishing mortality^{1,5,25}. This is supported by our analysis, in which declining trends in average age or length through time are seen for all exploited species (Fig. 3 in Supplementary Data). Reducing the average age and length of individuals within a population can increase recruitment variability by diminishing the capacity to weather short-term unfavourable environmental conditions. Many fish species use bet-hedging strategies to increase the survival rate of larvae under harsh and variable environmental conditions. Such hedging strategies are associated with

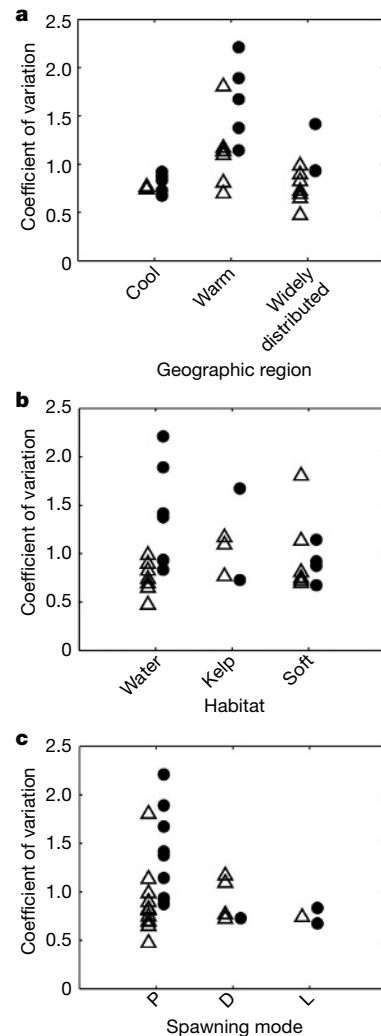


Figure 2 | Coefficients of variation of annual larval abundance of exploited and unexploited species associated with geographic regions, habitats and spawning modes. a–c, Filled circles indicate exploited species whereas open triangles indicate unexploited species associated with geographic regions (a), habitats (b) and spawning modes (c). P, pelagic spawners; D, demersal spawners; L, live-bearers. No significant association between coefficients of variation and habitats or spawning modes are found. For geographic regions (a), coefficients of variation for warm-water species are higher than those for cool-water species and widely distributed species. However, after accounting for the effect of regions, the coefficients of variation for the exploited species are still higher than those for the unexploited species (multiple regression, $P = 0.005$).

Table 1 | Results of the correlation analysis between fish abundance and time

	Species	Correlation coefficient	P-value
Exploited	<i>Engraulis mordax</i>	-0.126	0.439
Exploited	<i>Merluccius productus</i>	0.072	0.659
Exploited	<i>Sardinops sagax</i>	0.583	<0.001
Exploited	<i>Scomber japonicus</i>	0.159	0.327
Exploited	<i>Trachurus symmetricus</i>	-0.451	0.004*
Exploited	<i>Microstomus pacificus</i>	0.105	0.521
Exploited	<i>Paralabrax clathratus</i>	0.179	0.270
Exploited	<i>Paralichthys californicus</i>	-0.010	0.950
Exploited	<i>Parophrys vetulus</i>	-0.050	0.758
Exploited	<i>Scorpaenichthys marmoratus</i>	-0.097	0.550
Exploited	<i>Sebastes aurora</i>	-0.243	0.131
Exploited	<i>Sebastes paucispinis</i>	-0.505	0.001*
Exploited	<i>Sphyræna argentea</i>	0.230	0.154
Unexploited	<i>Cololabis saira</i>	0.037	0.822
Unexploited	<i>Ichthyos lockingtoni</i>	-0.569	<0.001*
Unexploited	<i>Leuroglossus stilbius</i>	-0.359	0.023*
Unexploited	<i>Tetragonurus cuvieri</i>	0.100	0.539
Unexploited	<i>Trachipterus altivelis</i>	0.040	0.809
Unexploited	<i>Chromis punctipinnis</i>	-0.099	0.544
Unexploited	<i>Lyopsetta exilis</i>	0.172	0.289
Unexploited	<i>Hippoglossina stomata</i>	-0.097	0.554
Unexploited	<i>Pleuronichthys verticalis</i>	0.069	0.673
Unexploited	<i>Sebastes jordani</i>	-0.206	0.203
Unexploited	<i>Symphurus atricaudus</i>	-0.123	0.448
Unexploited	<i>Zaniolepis frenata</i>	0.541	<0.001
Unexploited	<i>Argentina sialis</i>	0.437	0.005
Unexploited	<i>Hypsoblennius jenkinsi</i>	0.042	0.799
Unexploited	<i>Ophiodon scrippsae</i>	-0.167	0.303
Unexploited	<i>Oxylebius pictus</i>	0.189	0.244

Exploited species are defined as fisheries-targeted species⁴. Asterisks indicate a significant negative correlation between fish abundance and time. Only four among the 29 species show significant declining trends. No systematic differences exist between the exploited and unexploited groups in the prevalence of declining trends.

long-tailed age structures (a long tail of old individuals in the age distribution), and include: age-related differences in spawning locations and time^{26,27}, and increased quantity and quality of eggs produced by older (experienced) or larger fish²⁸. A long-tailed age structure can dampen environmental stochasticity and thus stabilize fish populations. In contrast, when fishing truncates age structure, fish populations become more variable because bet-hedging strategies are undermined and the populations more closely track short-term environmental variability. This well-documented mechanism^{26–28} suggests how fishing can make populations more susceptible to extrinsic environmental forcing. We call this phenomenon the age truncation effect (the ATE phenomenon), and show it is an effect that can operate independently of significant declining trends in abundances.

To our knowledge, our results provide the first empirical evidence to show that fishing increases variability in the abundance of exploited populations (even after accounting for life-history effects, ecological traits, phylogeny and a changing environment). The elevated variability of exploited populations is probably the result of the increased importance of recruitment and the elevated variability of recruitment caused by fishery-induced truncation of age structure^{1,5,8,9} (the ATE phenomenon). Indeed, as has been speculated²⁹, it is quite plausible that at low stock sizes the only regulatory process operating is a stochastic one. The most immediate implication for the management of fisheries is that, beyond the potential for causing a decline in abundance, fishing can provoke greater variability in exploited populations (and therefore reduced resilience) and thereby increase the risk of collapse of a fishery from stochastic environmental events¹¹. Obviously, this risk increases if fishing results in both higher variability and declining populations. That these two undesirable consequences of fishing can occur together represents a double jeopardy and should be of concern to fisheries managers.

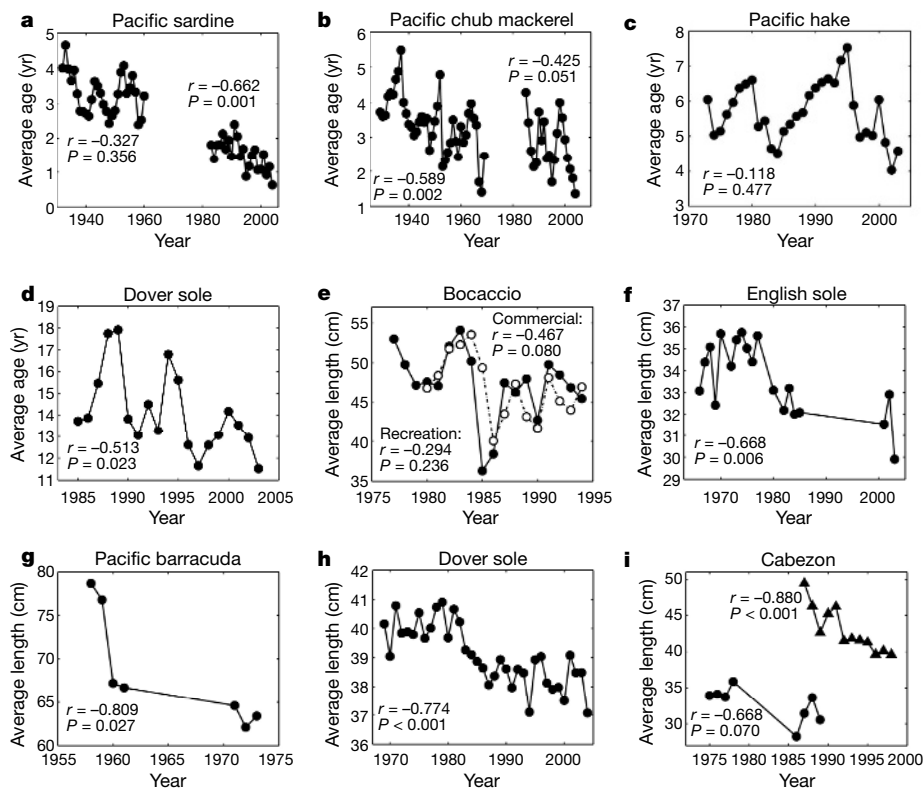


Figure 3 | Long-term declining trends in the average age and length of exploited species. a–i, Data are from fisheries catch records (see Supplementary Data). In a and b, analysis is carried out separately in the period before and after a moratorium on fishing. In e, filled and open circles represent data from recreational and commercial fisheries, respectively. In

i, circles and triangles represent data from southern and northern California, respectively. The P-values are corrected for serial dependence in the time series. Although only 7 out of 13 time series show a significant trend at $P < 0.05$, all exhibit a declining trend, which is highly significant as an ensemble result (binomial test, $P < 0.005$).

METHODS

The larval fish data used here were collected in the California Cooperative Oceanic Fisheries Investigations (CalCOFI). The CalCOFI programme is one of the most comprehensive observational oceanography programmes in the world with at least four cruises per year (except for tri-annual cruises between 1966 and 1984) and 66 stations per cruise, from 1949 (ref. 4). This data set provides fishery-independent data that are free from confounding effects (changes in fishing gear or areas) commonly associated with fishery catch data in estimating fish abundances. Because the CalCOFI programme spans more than 50 yr, the data reflect how fish populations respond at various time scales (from annual to decadal) of environmental forcing, as well as how they react to fishing. Because the fish populations live in the same area, they experience much the same large-scale environmental forcing. Importantly, both exploited and unexploited species were consistently sampled. These properties allow us to separate fishing effects from environmental effects on fish dynamics. We can use unexploited species as an objective reference and regard fishing as a treatment in a long-term experiment.

In this study, we used the well-documented assumption that larval fish abundances are proportional to the standing stock of the adults that produced them⁴. Evidence for this comes from studies that demonstrate correlations between larval counts and estimates of adult biomass (obtained independently from other surveys and fisheries assessment models)⁴. In addition to responding to changes in adult biomass, in our study counts of larvae may potentially vary due to changes in the reproductive effort of adults, mortality rates of eggs and early larvae, and large-scale movements of reproducing adults in and out of the CalCOFI survey grid. Despite these potential sources of added variability, the larval data for most species in our study were shown to track faithfully long-term variation in adult biomass (see Supplementary Data and Supplementary Table 1).

We used coefficients of variation of the annual abundance of larval fish as the measure of temporal variability for fish species. The coefficient of variation is a useful measurement of temporal variability because it is unitless (and therefore is suitable for cross-species comparison²⁰) and because each time series contains the same sample size²⁰. Our aim was to test whether the coefficients of variation of the exploited group were higher than those of the unexploited group, after factoring out the effects of ancillary variables (life-history traits, larval abundance, ecological traits and phylogeny). Stepwise multiple linear regression analysis was used to test the effect of fishing (dummy variable) and the ancillary variables on the coefficients of variation as well as the interactions between fishing and the ancillary variables on the coefficients of variation. Because of limited sample sizes we implemented forward stepwise regression to select sequentially significant variables to incorporate in the regression model. We used an *F*-test at each step to identify the variable with the most significant *P*-value for inclusion in the next step. We assumed that there are no systematic differences between exploited and unexploited populations with regard to inter-specific interactions. By-catch effects on unexploited species were assumed to be minimal⁴.

Missing data on life-history traits in the multiple regression were accounted for using the multiple imputation method³⁰ (1,000 imputations). Because life-history traits are correlated²², we used their correlation structure to impute the missing data, assuming a multivariate normal distribution³⁰. The statistical inferences were generated by combining the results of the 1,000 analyses³⁰.

Received 2 May; accepted 4 September 2006.

1. Berkeley, S. A., Hixon, M. A., Larson, R. J. & Love, M. S. Fisheries sustainability via protection of age structure and spatial distribution of fish populations. *Fisheries* **29**, 23–32 (2004).
2. Daan, N., Christensen, V. & Cury, P. M. Quantitative ecosystem indicators for fisheries management. *ICES J. Mar. Sci.* **62**, 307–614 (2005).
3. Pikitch, E. K. *et al.* Ecosystem-based fishery management. *Science* **305**, 346–347 (2004).
4. Hsieh, C. H. *et al.* A comparison of long-term trends and variability in populations of larvae of exploited and unexploited fishes in the Southern California region: a community approach. *Prog. Oceanogr.* **67**, 160–185 (2005).
5. Hutchings, J. A. & Reynolds, J. D. Marine fish population collapses: consequences for recovery and extinction risk. *Bioscience* **13**, 297–309 (2004).
6. Garcia, S. M., Zerbi, A., Aliaume, C., Do Chi, T. & Lasserre, G. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. *FAO Fish. Tech. Pap.* **443**, 1–71 (2003).
7. Beddington, J. R. & May, R. M. Harvesting natural populations in a randomly fluctuating environment. *Science* **197**, 463–465 (1977).

8. Murawski, S. A., Rago, P. J. & Trippel, E. A. Impacts of demographic variation in spawning characteristics on reference points for fishery management. *ICES J. Mar. Sci.* **58**, 1002–1014 (2001).
9. Murphy, G. I. Vital statistics of the Pacific sardine (*Sardinops Caerulea*) and the population consequences. *Ecology* **48**, 731–736 (1967).
10. Hsieh, C. H., Glaser, S. M., Lucas, A. J. & Sugihara, G. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* **435**, 336–340 (2005).
11. Scheffer, M., Carpenter, S., Foley, J. A., Folkes, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
12. Food and Agriculture Organization of the United Nations. (FAO). *Review of the State of World Marine Fishery Resources* (FAO, Rome, 2005).
13. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).
14. Hutchings, J. A. Collapse and recovery of marine fishes. *Nature* **406**, 882–885 (2000).
15. Beauprand, G., Brander, K. M., Lindley, J. A., Souissi, S. & Reid, P. C. Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661–664 (2003).
16. Finney, B. P., Gregory-Eaves, I., Douglas, M. S. V. & Smol, J. P. Fisheries productivity in the northeastern Pacific Ocean over the past 2,000 years. *Nature* **416**, 729–733 (2002).
17. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
18. May, R., Beddington, J. R., Horwood, J. W. & Shepherd, J. G. Exploiting natural populations in an uncertain world. *Math. Biosci.* **42**, 219–252 (1978).
19. Hilborn, R., Maquire, J., Parma, A. M. & Rosenberg, A. A. The precautionary approach and risk management: can they increase the probability of successes in fishery management? *Can. J. Fish. Aquat. Sci.* **58**, 99–107 (2001).
20. Pimm, S. L. *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities* 39–58 (University of Chicago, Chicago, 1991).
21. Ludwig, D., Hilborn, R. & Waters, C. Uncertainty, resource exploitation, and conservation: lessons from history. *Science* **260**, 17–36 (1993).
22. Winemiller, K. O. & Rose, K. A. Patterns of life-history diversification in North American fishes: implications for population regulation. *Can. J. Fish. Aquat. Sci.* **49**, 2196–2218 (1992).
23. Mertz, G. & Myers, R. A. Match/mismatch predictions of spawning duration versus recruitment variability. *Fish. Oceanogr.* **3**, 1–14 (1994).
24. Mertz, G. & Myers, R. A. Influence of fecundity on recruitment variability of marine fish. *Can. J. Fish. Aquat. Sci.* **53**, 1618–1625 (1996).
25. Hutchings, J. & Baum, J. Measuring marine fishes biodiversity: temporal changes in abundance, life history and demography. *Phil. Trans. R. Soc. Lond. B* **360**, 315–338 (2005).
26. Hutchings, J. A. & Myers, R. A. Effect of age on the seasonality of maturation and spawning of Atlantic cod, *Gadus morhua*, in the northwest Atlantic. *Can. J. Fish. Aquat. Sci.* **50**, 2468–2474 (1993).
27. Lambert, T. C. Duration and intensity of spawning in herring *Clupea harengus* as related to the age structure of the population. *Mar. Ecol. Prog. Ser.* **39**, 209–220 (1987).
28. Marteinsdottir, G. & Steinarrson, A. Maternal influence on the size and viability of Iceland cod (*Gadus morhua*) eggs and larvae. *J. Fish Biol.* **52**, 1241–1258 (1998).
29. Shepherd, J. G. & Cushing, D. H. Regulation in fish populations: myth or mirage? *Phil. Trans. R. Soc. Lond. B* **330**, 151–164 (1990).
30. Little, R. J. A. & Rubin, D. B. *Statistical Analysis with Missing Data* 164–220 (John Wiley & Sons, New York, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the CalCOFI for the use of the data. R. Hewitt, D. Checkley, M. Ohman, S. Sandin, M. Maunder, P. Hastings, K. Roy, D. Lucas and S. Glaser provided discussion and comments on this work. A. MacCall, K. Hill, D. Sampson, I. Stewart, J. Cope, D. Aseltin-Nelson and J. Mason provided fisheries-related data. B. Macewicz, N. Lo, B. Watson and S. Suyama provided data and comments on life-history traits of fishes. Our study was funded by NOAA Fisheries and the Environment (FATE), the National Marine Fisheries Service, NSF/LTER CCE 'Nonlinear Transitions in the California Current Coastal Pelagic Ecosystem', the Center for Marine Bioscience and Biotechnology, National Taiwan Oceanic University, the Deutsche Bank Complexity Studies Fund and the Sugihara Family Trust.

Author Contributions C.H. and G.S. conceived the original study. C.H. is responsible for the statistical analyses and uncovering the main result. All co-authors contributed to refining the analysis, framing and interpreting the result, and to its final exposition.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.S. (gsugihara@ucsd.edu).

An early evolutionary origin for the minor spliceosome

Anthony G. Russell^{1*}, J. Michael Charette^{1*}, David F. Spencer¹ & Michael W. Gray¹

The minor spliceosome is a ribonucleoprotein complex that catalyses the removal of an atypical class of spliceosomal introns (U12-type) from eukaryotic messenger RNAs^{1,2}. It was first identified and characterized in animals, where it was found to contain several unique RNA constituents that share structural similarity with and seem to be functionally analogous to the small nuclear RNAs (snRNAs) contained in the major spliceosome^{3–8}. Subsequently, minor spliceosomal components and U12-type introns have been found in plants^{9–12} but not in fungi. Unlike that of the major spliceosome, which arose early in the eukaryotic lineage^{13,14}, the evolutionary history of the minor spliceosome is unclear because there is evidence of it in so few organisms. Here we report the identification of homologues of minor-spliceosome-specific proteins and snRNAs, and U12-type introns, in distantly related eukaryotic microbes (protists) and in a fungus (*Rhizopus oryzae*). Cumulatively, our results indicate that the minor spliceosome had an early origin: several of its characteristic constituents are present in representative organisms from all eukaryotic supergroups for which there is any substantial genome sequence information. In addition, our results reveal marked evolutionary conservation of functionally important sequence elements contained within U12-type introns and snRNAs.

In a comparison of expressed sequence tag (EST) data from an amoeboid protist (*Acanthamoeba castellanii*; Fig. 1) to the corresponding genome sequence, we discovered an unusual intron in a gene encoding a COMMD-domain-containing protein (COMMD7). The EST (GenBank accession code EC110939) confirms that the intron boundary nucleotides are 'AT...AC' (Supplementary Fig. S1A), rather than the 'GT...AG' that is characteristic of most spliceosomal introns. At both its 5' and 3' splice sites and at its predicted branch-point sequence, the *Acanthamoeba* intron shares substantial sequence identity with conserved sequence elements diagnostic of U12-type minor spliceosomal introns^{1,2,15,16} (Fig. 2a).

Because this represented the first example of a minor spliceosomal intron in a unicellular eukaryote, we searched the partial *A. castellanii* genome sequence¹⁷ for additional U12-type introns. Using the pattern search algorithm PatScan¹⁸, and allowing for known sequence degeneracy in the conserved elements of characterized animal and plant U12-type introns, we identified four additional minor spliceosomal intron candidates in *Acanthamoeba* (Fig. 2a). Two of the *A. castellanii* introns are predicted to have the 'GT...AG' boundary dinucleotides more frequently observed in minor spliceosomal introns in humans and plants (*Arabidopsis*)^{9,12,19}. The predicted branch-point adenosines of the three 'AT...AC' or 'AT...AA' introns are only 9–11 nucleotides from the intron 3' ends, a property consistent with similar distance constraints (10–16 nucleotides) seen in most members of the combined set of human and plant 'AT...AC' introns^{12,20}.

An *A. castellanii* EST (GenBank EC103018) confirms the removal and predicted splicing boundaries of the intron from the 'RNA processing factor 1' gene transcript (Supplementary Fig. S1B). The intron in the 'serine/threonine kinase 11' gene is unusual because of its predicted 'AA' 3' boundary sequence (Supplementary Fig. S1C); however, this sequence functions as an efficient 3' splice site for mammalian U12-type introns²¹. A fourth intron is located in an open reading frame (*Aca* ORF 1) encoding a protein of unknown function also found in the amoeboid protist *Dictyostelium discoideum* (Supplementary Fig. S1D).

In the absence of EST data for the latter two introns, polymerase chain reaction with reverse transcription (RT-PCR) was used to verify the removal of the introns and to define their boundaries. Complementary DNAs corresponding to spliced mRNAs of the expected size were detected (Fig. 2b), cloned and sequenced, as were PCR products generated for the respective amplified genes. The

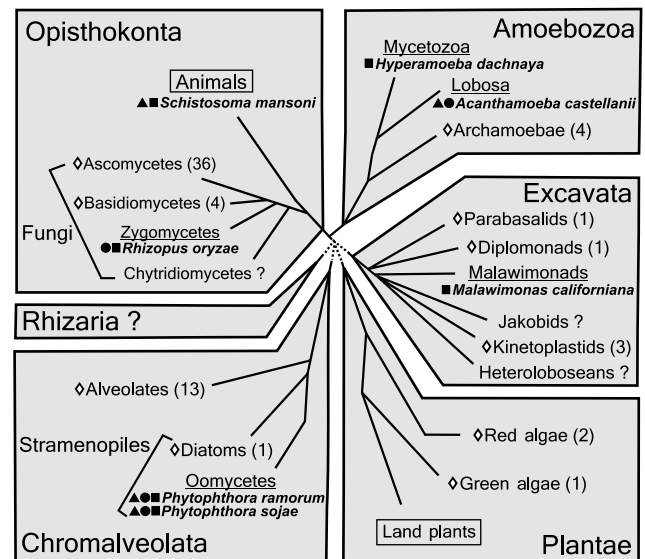


Figure 1 | Phylogenetic distribution of eukaryotic organisms that contain minor spliceosomal components. Shown are the phylogenetic positions of organisms in which minor spliceosomal proteins (filled squares), snRNAs (filled triangles) and/or introns (filled circles) were identified in this study. Also shown are groups and the number of species (parentheses) for which substantial nuclear genome information is available but in which minor spliceosomal components are not evident (open diamonds). Groups previously known to contain these components (animals and land plants) are boxed. Question marks indicate groups of particular interest for which significant genomic sequence information is not yet available. This representation, within which the relative branching order among six eukaryotic supergroups is uncertain (dotted lines), is adapted from ref. 30.

¹Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia B3H 1X5, Canada.

*These authors contributed equally to this work.

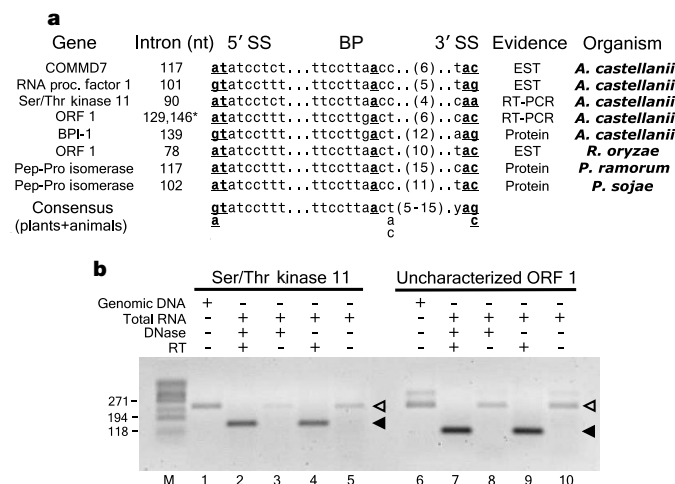


Figure 2 | U12-type minor spliceosomal introns in unicellular eukaryotes. **a**, Intron features, sizes and gene locations. Shown are intron sequences that flank the splice sites (SS) and branch points (BP) and that contain the conserved boundary dinucleotides and branch-point adenosine (bold and underlined). Numbers in parentheses specify additional nucleotides. Two copies of the *Aca* ORF 1 gene contain different-sized introns (asterisk). Less frequent nucleotides at select positions are indicated below the consensus sequence¹. **b**, Experimental verification of *A. castellanii* introns. RT-PCR was carried out using total RNA either with (lanes 2, 3, 7 and 8) or without (4, 5, 9 and 10) pre-treatment with DNase I. Filled arrowheads indicate cDNA products (resolved on a 1% agarose gel) corresponding to spliced mRNAs. Sizes of select bands of a DNA marker (M) are indicated. PCR reactions with genomic DNA (lanes 1 and 6) show the sizes of the intron-containing products (open arrowheads).

results confirmed the sequences, removal of the introns, and splicing boundaries. For *Aca* ORF 1, an additional gene copy containing a larger 'AT...AC' U12-type intron at the identical position was also detected and characterized (Supplementary Fig. S1D). The fifth U12-type intron in *A. castellanii* is located in a BPI-1 (bactericidal permeability increasing) gene encoding a protein with orthologues in other eukaryotes (Supplementary Fig. S1E). These combined results indicate that *A. castellanii* may possess a minor spliceosome that removes these atypical introns.

In addition to protein constituents that it shares with the major spliceosome²², the minor spliceosome contains specific proteins that have been identified as components of the U11–U12 di-snRNP particle^{22,23}. Using BLAST analysis, we searched available eukaryotic genome data for sequences encoding homologues of these proteins, focusing primarily on unicellular organisms. We also searched EST sequence data available through the Protist EST Program (PEP; http://www.bch.umontreal.ca/pepdb/pep_main.html). We identified predicted orthologues for six of the seven characterized human minor-spliceosome-specific proteins: namely, the 20K, 25K, 31K (also known as MADP-1), 35K, 59K and 65K proteins (Table 1 and Supplementary Fig. S2). The identified set includes orthologues in

representative genera (*Malawimonas* and *Phytophthora*) of two of the eukaryotic supergroups (Excavata and Chromalveolata, respectively; Fig. 1) not previously known to contain minor spliceosomes or U12-type introns. In addition, we identified orthologues in other unanticipated taxa, including *Rhizopus oryzae* (representing, to our knowledge, the first known minor spliceosomal components in fungi) and *Schistosoma mansoni*, a trematode. We also searched the partial genome data (estimated at ~0.5× coverage¹⁷) of *A. castellanii*. So far we have not identified evident homologues of these proteins in this organism, a result that is most probably due to the incomplete nature of this data set (see below).

Of note, excepting only *Hyperamoeba dachnaya*, we identified several minor-spliceosome-specific proteins in the same organism (Table 1), strongly indicating that these taxa contain minor spliceosomes. Because much of the available genomic information is still incomplete, our results do not represent a complete account of the occurrence of these proteins in the minor spliceosomes of these organisms, or in eukaryotes in general. Furthermore, for some of the organisms examined here (particularly Excavates), little or no genomic information and only limited EST sequence data are currently available. Consequently, our analysis is likely to underestimate the number of organisms that contain minor spliceosomes.

Having obtained evidence for minor spliceosomes in *R. oryzae* and two *Phytophthora* species, we searched for U12-type introns in available nuclear genome sequence data. In *R. oryzae*, we identified two intron candidates containing 'AT...AC' boundaries and the expected conserved elements of U12-type introns. The first intron is located in a gene of unassigned function (*Ror* ORF 1; Fig. 2a and Supplementary Fig. S3A), for which an *R. oryzae* EST sequence (EE009306) confirms its removal and precise splicing boundaries. The second intron is present in a 'DNA-activated, protein kinase-like' gene encoding a conserved protein (Supplementary Fig. S3). In the two *Phytophthora* species, *P. ramorum* and *P. sojae*, U12-type introns (conserved between the two species) were identified in two different genes. The first intron is located in a gene encoding a highly conserved 'FKBP-type peptidyl-prolyl *cis-trans* isomerase' protein (Fig. 2a and Supplementary Fig. S3). In both species, removal of the intron from the precursor mRNA is required to maintain amino-acid conservation of the translated protein downstream of the intron insertion site. The second intron is located in a gene that encodes a protein of unassigned function that is conserved between the two species (Supplementary Fig. S3).

In summary, we predicted which organisms are likely to contain U12-type introns on the basis of the identification of minor spliceosomal protein orthologues. Subsequently, we identified U12-type introns in very distantly related eukaryotes on the basis of the marked sequence conservation of both their 5' splice boundaries and branch-point sequences. This conservation further implies that these phylogenetically disparate taxa use very similar splicing mechanisms in the removal of these introns.

In a parallel study, we identified sequences encoding snRNA homologues from the *A. castellanii* genome data, including a major

Table 1 | Identification of orthologues of minor-spliceosome-specific proteins

Organism	Supergroup*	Subdivision	Protein					
			20K	25K	31K†	35K‡	59K	65K‡
<i>Schistosoma mansoni</i>	Opisthokonta	Trematodes	Yes§		Yes§			Yes§
<i>Rhizopus oryzae</i>	Opisthokonta	Fungi	Yes§			Yes§	Yes§	Yes§
<i>Hyperamoeba dachnaya</i>	Amoebozoa	Mycetozoa			Yes			
<i>Malawimonas californiana</i>	Excavata	Malawimonads		Yes	Yes			
<i>Phytophthora ramorum</i>	Chromalveolata	Stramenopiles			Yes§	Yes§		Yes§
<i>Phytophthora sojae</i>	Chromalveolata	Stramenopiles			Yes§	Yes§		Yes§

* Supergroup classification is from ref. 30.

† Also referred to as MADP-1.

‡ Sequence comparisons (Supplementary Fig. S2) and phylogenetic reconstructions (data not shown) effectively distinguish the 35K and 65K minor spliceosome proteins from sequence-related major spliceosome components (U1-70K and U1A/U2B*, respectively).

§ Amino-acid sequence derived from genomic DNA information. 'Yes' indicates identification of a homologue.

|| Amino-acid sequence derived from EST information.

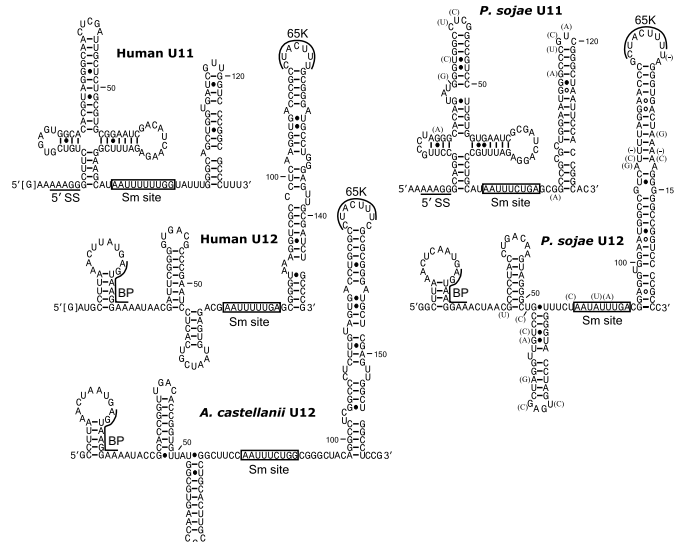


Figure 3 | Secondary structures of minor spliceosomal U11 and U12 snRNA homologues. Nucleotides predicted to interact with the 5' splice sites (5' SS) or branch points (BP) of U12-type introns are under- or overlined, respectively. Known or predicted Sm-protein-binding sites are boxed. In U12 snRNA, a highly conserved sequence element recognized by the human minor-spliceosome-specific 65K protein²⁶ is highlighted. The human structures are adapted from refs 1, 2, and 26. Trimethylated 5' guanosine caps are indicated in square brackets. For *P. sojae*, nucleotide substitutions or deletions at corresponding positions in the *P. ramorum* RNAs are indicated in parentheses. The *P. ramorum* U12 sequence lacks nucleotides comprising the predicted *P. sojae* U114–A145 basepair. The 5' and 3' ends of these RNAs are approximations based on sequence conservation between the two species.

spliceosomal U2 snRNA variant (Supplementary Fig. S4). More germane to this study, we have also identified a putative minor spliceosomal U12 snRNA, a functional analogue of U2 snRNA (Fig. 3). This *Acanthamoeba* U12 snRNA has significant structural and sequence similarity to its human counterpart, particularly in the 5' stem-loop structure. The latter structure contains the sequence that interacts with the branch-point sequence of U12-type introns, and its perfect conservation between the human and *A. castellanii* RNAs explains the observed conservation of branch-point sequences in the identified *A. castellanii* U12-type introns. To confirm the expression of this *Acanthamoeba* U12 snRNA gene and to define accurately the ends of the mature RNA, we carried out both 5' and 3' rapid amplification of cDNA ends (RACE; Supplementary Figs S4B and S5A).

On the basis of the conservation of sequence elements in the newly identified U12-type introns and in the *Acanthamoeba*, human and plant snRNAs, we assigned parameters for a search of genomic sequences encoding minor spliceosomal U11 and U12 snRNAs in the *Phytophthora* species. Minor spliceosomal snRNAs can be difficult to identify by bioinformatic approaches because of limited sequence conservation and, in some cases (such as *Drosophila*)^{24,25}, substantial structural divergence. Nevertheless, we reasoned that the *Phytophthora* snRNAs should possess sequences capable of base-pairing to the conserved elements in the newly identified introns and should also show sequence and structural conservation between the two species. Using PatScan, we identified predicted homologues of the U11 and U12 snRNAs in *P. ramorum* and *P. sojae* (Fig. 3). Compensatory nucleotide substitutions in the predicted RNA secondary structures indicate structural conservation between the two species. The sequence elements predicted to base-pair to the introns are in structural environments identical to those seen in the human, plant and *A. castellanii* snRNAs. Further evidence supporting our inference that these are snRNA homologues derives from predicted Sm-protein-binding sites, the location of the encoding sequences in intergenic regions and the confinement of near-perfect sequence

identity between the two species to the genomic regions encoding and immediately flanking the mature snRNA sequences. The perfect conservation of a sequence element in the extended 3' stem-loop structure of the U12 snRNAs among these very distantly related eukaryotes is striking. This region of the RNA has been shown to be a recognition site for the binding of the minor spliceosomal 65K protein²⁶.

Our identification of U11 and U12 snRNA homologues, minor-spliceosome-specific proteins and U12-type introns is compelling evidence for the existence of minor spliceosomes in very diverse, distantly related, unicellular eukaryotes. The distribution of these components within the eukaryotes is noteworthy in that there is now evidence for their existence in organisms situated immediately basal to (or even located within) eukaryotic groups that seem to (or predominantly) lack them completely. This distribution suggests that there has been widespread, independent purging of minor spliceosomal components throughout eukaryotes during their evolutionary diversification. In Fungi, the asco- and basidiomycetes seem to lack minor spliceosome components but *Rhizopus* (Zygomycota) has both the proteins and introns. Notably, the two *Rhizopus* genes containing U12-type introns do not have evident homologues in these other fungi. Frequent (or early) loss may also have occurred in the alveolates in the supergroup Chromalveolata. These observations support an earlier suggestion that U12-type introns were once more prevalent in eukaryotes⁹. The possibility remains that other, more divergent, minor spliceosomal systems have escaped detection; however, the discovery of such systems would only further emphasize the ancient nature of this spliceosome.

Our results raise additional questions about the evolutionary relationship of the major and minor spliceosomes. Shared components (proteins and the U5 snRNA) and the structural similarity of functionally analogous but distinct snRNAs suggest an early common origin—a proposal that is made more plausible by our finding of minor spliceosomal components throughout the eukaryotic radiation. In this regard, the spliceosomal introns identified in the divergent protists *Trichomonas vaginalis*²⁷ (parabasalid) and *Giardia lamblia*²⁸ (diplomonad) are particularly intriguing. Similar to U12-type introns, these introns show extensive sequence conservation at both their 5' boundaries and branch points, and their branch points are very close to the intron 3' ends; however, the elements conserved in these introns are not close sequence matches to the elements conserved in U12-type introns. In addition, we do not have clear evidence for the presence of any of the currently known minor-spliceosome-specific proteins in these organisms, raising the question of which class of spliceosomal intron the *Giardia* and *Trichomonas* ones represent. Characterization of the components and mechanism of the splicing systems in these organisms will undoubtedly help to answer this question.

METHODS

Minor spliceosomal intron, protein and snRNA searches. U12-type introns were identified using PatScan¹⁸ with the parameters specified in Supplementary Fig. S5B. Positive hits were screened for location in a gene encoding an ORF and for whether or not the corresponding ORF is conserved in other eukaryotic taxa. For *A. castellanii* and *R. oryzae*, positive hits were also compared, by both BLASTN and TBLASTX, to the complete compilation of EST sequences in the Protist EST database, TBestDB (<http://amoebidia.bcm.umontreal.ca/pepdb/searches/login.php>).

The *A. castellanii* and *S. mansoni* U12 snRNA genes were identified by BLASTN searches against the partial genome data for these organisms by using the human U12 sequence as query. The *Phytophthora* U12 genes were identified with PatScan and the search parameters [CCTTAACT(1,0,0)...(2–5 nucleotides)...AGTAAGG(1,0,0)]. For every positive hit obtained for a species, 500 nucleotides of additional upstream and downstream flanking genomic sequence were also compared by BLASTN to the genome of the other species. For the surviving hits showing extended sequence conservation between species beyond the short sequence strings identified by PatScan, structural potential was assessed by using known U12 snRNA structures as guides. The *Phytophthora*

U11 genes were identified by a similar but more complex search strategy (Supplementary Fig. S5C). Genomic sequences encoding U2 snRNAs were identified by BLASTN analysis using the conserved 5' half of the human U2 snRNA sequence as query.

Minor-spliceosome-specific protein homologues were identified by TBLASTN searches against genomic and EST data using the human, plant and newly identified protist protein sequences as queries.

RACE analysis and intron verification. Both 5' and 3' RACE analyses were done as described²⁹ to map the extremities of the *A. castellanii* U12 snRNA. The RT-PCR experiments used to confirm the *A. castellanii* U12-type introns were done as described²⁸. Oligonucleotides for these experiments are given in Supplementary Fig. S5A.

Received 13 July 2006; accepted 31 August 2006.

- Patel, A. A. & Steitz, J. A. Splicing double: insights from the second spliceosome. *Nature Rev. Mol. Cell Biol.* **4**, 960–970 (2003).
- Will, C. L. & Lührmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.* **386**, 713–724 (2005).
- Tarn, W.-Y., Yario, T. A. & Steitz, J. A. U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* **1**, 644–656 (1995).
- Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**, 1716–1718 (1996).
- Tarn, W.-Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT–AC) intron *in vitro*. *Cell* **84**, 801–811 (1996).
- Tarn, W.-Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT–AC introns. *Science* **273**, 1824–1832 (1996).
- Yu, Y.-T. & Steitz, J. A. Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT–AC intron. *Proc. Natl Acad. Sci. USA* **94**, 6030–6035 (1997).
- Kolossova, I. & Padgett, R. A. U11 snRNA interacts *in vivo* with the 5' splice site of U12-dependent (AU–AC) pre-mRNA introns. *RNA* **3**, 227–233 (1997).
- Burge, C. B., Padgett, R. A. & Sharp, P. A. Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**, 773–785 (1998).
- Shukla, G. C. & Padgett, R. A. Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* **5**, 525–538 (1999).
- Lorković, Z. J., Lehner, R., Forstner, C. & Barta, A. Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA* **11**, 1095–1107 (2005).
- Zhu, W. & Brendel, V. Identification, characterization and molecular phylogeny of U12-dependent introns in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* **31**, 4561–4572 (2003).
- Collins, L. & Penny, D. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* **22**, 1053–1066 (2005).
- Rodríguez-Trelles, F., Tarrío, R. & Ayala, F. J. Origin and evolution of spliceosomal introns. *Annu. Rev. Genet.* doi:10.1146/annurev.genet.40.110405.090625 (published online 12 June 2006).
- Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357–365 (1994).
- Dietrich, R. C., Incorvaia, R. & Padgett, R. A. Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* **1**, 151–160 (1997).
- Anderson, I. J. *et al.* Gene discovery in the *Acanthamoeba castellanii* genome. *Protist* **156**, 203–214 (2005).
- Dsouza, M., Larsen, N. & Overbeek, R. Searching for patterns in genomic data. *Trends Genet.* **13**, 497–498 (1997).
- Levine, A. & Durbin, R. A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.* **29**, 4006–4013 (2001).
- Dietrich, R. C., Peris, M. J., Seyboldt, A. S. & Padgett, R. A. Role of the 3' splice site in U12-dependent intron splicing. *Mol. Cell. Biol.* **21**, 1942–1952 (2001).
- Dietrich, R. C., Fuller, J. D. & Padgett, R. A. A mutational analysis of U12-dependent splice site dinucleotides. *RNA* **11**, 1430–1440 (2005).
- Will, C. L., Schneider, C., Reed, R. & Lührmann, R. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* **284**, 2003–2005 (1999).
- Will, C. L. *et al.* The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* **10**, 929–941 (2004).
- Schneider, C., Will, C. L., Brosius, J., Frilander, M. J. & Lührmann, R. Identification of an evolutionarily divergent U11 small nuclear ribonucleoprotein particle in *Drosophila*. *Proc. Natl Acad. Sci. USA* **101**, 9584–9589 (2004).
- Otake, L. R., Scamborova, P., Hashimoto, C. & Steitz, J. A. The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol. Cell* **9**, 439–446 (2002).
- Benecke, H., Lührmann, R. & Will, C. L. The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11–59K protein. *EMBO J.* **24**, 3057–3069 (2005).
- Vaňáčková, S., Yan, W., Carlton, J. M. & Johnson, P. J. Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **102**, 4430–4435 (2005).
- Russell, A. G., Shutt, T. E., Watkins, R. F. & Gray, M. W. An ancient spliceosomal intron in the ribosomal protein L7a gene (*Rpl7a*) of *Giardia lamblia*. *BMC Evol. Biol.* **5**, 45 (2005).
- Russell, A. G., Schnare, M. N. & Gray, M. W. A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: structural, functional and evolutionary insights. *J. Mol. Biol.* **357**, 1548–1565 (2006).
- Keeling, P. J. *et al.* The tree of eukaryotes. *Trends Ecol. Evol.* **20**, 670–676 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. N. Schnare, T. E. Shutt and A. J. Lohan for discussions; M. Dłutek for automated DNA sequencing; and the various genome sequencing centres for making preliminary data publicly available, including the Broad Institute of Harvard and MIT (*Rhizopus*) and the Joint Genome Institute (*Phytophthora*). This work was supported by funding (to M.W.G.) from the Canadian Institutes of Health Research and Genome Canada (Protist EST Program). M.W.G. acknowledges salary support from the Canada Research Chairs Program and the Canadian Institute for Advanced Research.

Author Contributions J.M.C. initially discovered the first protist minor-spliceosomal snRNA (*A. castellanii* U12) and A.G.R. later identified the *Phytophthora* U11 and U12 snRNAs. A.G.R. did the intron and protein orthologue searches and experimentally verified intron removal and snRNA gene expression. D.F.S. assisted with computer analysis and isolated *A. castellanii* nucleic acids. M.W.G. discovered the COMMD7 intron and initiated a search for minor-spliceosome-specific protein orthologues in protists. A.G.R. wrote the paper. All authors discussed the results and critiqued the final manuscript.

Author Information The *A. castellanii* U12 snRNA sequence and the additional gene copy of *Aca* ORF 1 are deposited in Genbank (DQ888370 and DQ888371, respectively). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W.G. (m.w.gray@dal.ca).

A linguistic model for the rational design of antimicrobial peptides

Christopher Loose^{1*}, Kyle Jensen^{1,2,3*}, Isidore Rigoutsos^{1,4} & Gregory Stephanopoulos¹

Antimicrobial peptides (AmPs) are small proteins that are used by the innate immune system to combat bacterial infection in multicellular eukaryotes¹. There is mounting evidence that these peptides are less susceptible to bacterial resistance than traditional antibiotics and could form the basis for a new class of therapeutic agents². Here we report the rational design of new AmPs that show limited homology to naturally occurring proteins but have strong bacteriostatic activity against several species of bacteria, including *Staphylococcus aureus* and *Bacillus anthracis*. These peptides were designed using a linguistic model of natural AmPs: we treated the amino-acid sequences of natural AmPs as a formal language and built a set of regular grammars to describe this language. We used this set of grammars to create new, unnatural AmP sequences. Our peptides conform to the formal syntax of natural antimicrobial peptides but populate a previously unexplored region of protein sequence space.

AmPs are ubiquitous among multicellular eukaryotes and are found in diverse contexts including frog skin, scorpion venom and human sweat. Recent studies show that some AmPs are active against pathogens that are resistant to traditional antibiotics such as penicillin, tetracycline and vancomycin³. In addition to their antibiotic uses, AmPs might have other interesting clinical applications: they act as adjuvants for the adaptive immune system⁴ and might be useful in treating certain cancers⁵.

The many disease-relevant actions of AmPs are a consequence of their broad ability to distinguish eukaryotic cells from pathogenic invaders. In general, AmPs have a net positive charge and an amphipathic three-dimensional structure that gives the peptides an electrostatic affinity to the outer leaflet of the microbial membrane⁶. This affinity leads to binding, disruption of the membrane and, ultimately, microbial cell death⁷.

Our preliminary studies of natural AmPs indicated that their amphipathic structure gives rise to a modularity among the different AmP amino-acid sequences. The repeated usage of sequence modules—which might be a relic of evolutionary divergence and radiation—is reminiscent of phrases in a natural language, such as English. For example, the pattern QxEGxLxKxxK (where 'x' is any amino acid) is found in more than 90% of the insect AmPs known as cecropins. On the basis of this observation, we modelled the AmP sequences as a formal language—a set of sentences using words from a fixed vocabulary. In this case, the vocabulary is the set of naturally occurring amino acids, represented by their one-letter symbols⁸.

We conjectured that the 'language of AmPs' could be described by a set of regular grammars. Regular grammars are, in essence, simple rules that describe the allowed arrangements of words. These grammars, such as the cecropin pattern mentioned previously,

are commonly written as regular expressions and are widely used to describe patterns in nucleotide and amino-acid sequences^{9,10}.

To find a set of regular grammars to describe AmPs we used the Teiresias pattern discovery tool¹¹. With Teiresias, we derived a set of 684 regular grammars that occur commonly in 526 well-characterized eukaryotic AmP sequences from the Antimicrobial Peptide Database (APD)¹² (see Methods). Together, these ~700 grammars describe the 'language' of the AmP sequences. In this linguistic metaphor, the peptide sequences are analogous to sentences and the individual amino acids are analogous to the words in a sentence. Each grammar describes a common arrangement of amino acids, similar to popular phrases in English. For example, the frog AmP brevinin-1E contains the amino-acid sequence fragment PKIFCKITRK, which matches the grammar P[KAYS][ILN][FGI]C[KPSA][IV][TS][RKC][KR] from our database (the bracketed expression [KAYS] indicates that, at the second position in the grammar, lysine, alanine, tyrosine or serine is equally acceptable). On the basis of this match, we would say that the brevinin-1E fragment is 'grammatical'.

By design, each grammar in this set of ~700 grammars is ten amino-acids long and is specific to AmPs—at least 80% of the matches for each grammar in Swiss-Prot/TrEMBL¹³ (the APD is a subset of Swiss-Prot/TrEMBL) are found in peptides annotated as AmPs.

To design unnatural AmPs, we combinatorially enumerated all grammatical sequences of length twenty (see Methods) in which each window of size ten in the 20-mers was matched by one of the ~700 grammars. We chose this length because we could easily chemically synthesize 20-mers and this length is close to the median length of AmPs in the APD. From this set, we removed any 20-mers that had six or more amino acids in a row in common with a naturally occurring AmP; we then clustered the remaining sequences on the basis of similarity, choosing 42 to synthesize and assay for antimicrobial activity. This process is illustrated in Fig. 1 and the designed sequences are shown as Supplementary Table 1.

For each of the 42 synthetic peptides, we also designed a shuffled sequence, in which the order of the amino acids was rearranged randomly so that the sequence did not match any grammars. These shuffled peptides had the same amino-acid composition as their synthetic counterparts and thus, the same molecular weight, charge and isoelectric point (bulk physicochemical factors that are often correlated with antimicrobial activity). We hypothesized that because the shuffled sequences were 'ungrammatical' they would have no antimicrobial activity, despite having the same bulk physicochemical characteristics. In addition, we selected eight peptides from the APD as positive controls and six 20-mers from non-antimicrobial proteins as negative controls. A complete list of designed and shuffled sequences, along with controls, is provided in the Supplementary Information.

¹Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Harvard-MIT Health Sciences and Technology, Cambridge, Massachusetts 02139, USA. ³Agrivida, 411 Massachusetts Ave B1, Cambridge, Massachusetts 02139, USA. ⁴IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA.

*These authors contributed equally to this work.

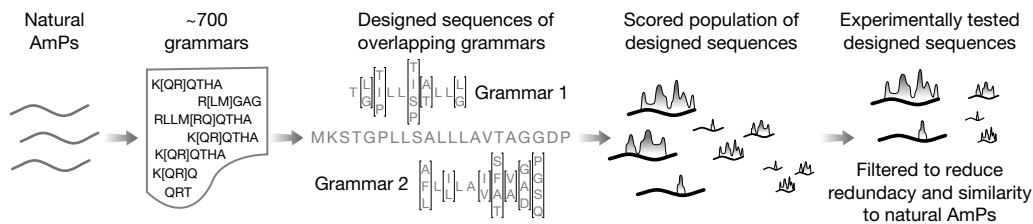


Figure 1 | A schematic of the *in silico* peptide design strategy. Grammars are induced from the set of natural AmP sequences using Teiresias. Overlapping grammars are stitched together to create new 20-amino-acid sequences that correspond to the antimicrobial syntax. Designed sequences

are scored on the basis of the prevalence of the underlying grammar in natural peptides. After removing designs with six or more consecutive amino acids in common with a natural peptide, a representative set of sequences is selected for experimental validation.

We characterized the activity of each synthetic AmP using a broth microdilution assay described elsewhere¹⁴. This assay measures the minimum inhibitory concentration (MIC) at which the peptide inhibits growth of the target organism. Table 1 shows the MICs of synthetic peptides against *Bacillus cereus* and *Escherichia coli*, as representative gram-positive and gram-negative bacteria. Of the 40 soluble (two of the designed and four of the shuffled peptides were insoluble) designed peptides, 18 had an MIC of 256 µg ml⁻¹ or less against at least one of the bacterial targets. Only two of the soluble, shuffled peptides showed antibacterial activity. Thus, the activity is not an artefact of molecular weight, charge or isoelectric point.

Of the negative controls—six peptides randomly selected from the middle of non-antimicrobial proteins from Swiss-Prot/TrEMBL—none had antimicrobial activity, whereas six of the eight naturally-occurring AmPs in the positive control group did.

To validate MIC determinations, we measured optical density at varying concentrations of a representative set of designed, shuffled, and natural peptides. Supplementary Fig. 1 shows that peptides inhibit growth at the MICs determined by visual inspection. In addition, plating of these samples confirmed that these peptide-treated bacterial populations were generally not viable at concentrations of peptides above the MIC.

Two of the designed peptides, D28 (FLGVVFKLASKVFPVFGKV) and D51 (FLFRVASKVFPALIGKFKKK), inhibited *B. cereus* growth at 16 µg ml⁻¹, which is close to the MICs of the strong positive controls melittin and cecropin-melittin hybrid (8 µg ml⁻¹). Peptides with activity against gram-positive bacteria are particularly exciting because of the prevalence of drug-resistant nosocomial *S. aureus* and the threat of bioterror agents such as *B. anthracis*, or anthrax. Therefore, we assayed seven designed peptides that had such activity, including the highly active D28 and D51 peptides, against the Smith diffuse strain of *S. aureus* and the Sterne strain of *B. anthracis*. As shown in Table 2, all seven peptides were active against both bacteria at 256 µg ml⁻¹ or less, whereas only one of the seven shuffled controls was active (the only active shuffled peptide had MICs of 128 µg ml⁻¹ against *S. aureus* and 256 µg ml⁻¹ against *B. anthracis*). Moreover, two designed peptides, D28 and D51, had MICs of 16 µg ml⁻¹ against *Bacillus anthracis*, which is equivalent to the activity of Cecropin-melittin hybrid, a strong natural AmP. D28 also had an MIC of 8 µg ml⁻¹ against *S. aureus*.

In an attempt to improve the antimicrobial activity of our designed peptides, we optimized our best candidate, peptide D28, using a heuristic approach. We created 44 variants of D28 by introducing mutations that were selected to increase positive charge, increase

hydrophobicity, remove an interior proline residue, or improve segregation of positive and hydrophobic residues based on a helical projection. As shown in Supplementary Table 2, 18 of the 44 D28 variants showed improved activity against *E. coli*, *B. cereus* or *S. aureus*. Many of the D28 variants with improved activity against *B. cereus* included a mutation at an internal proline, either to lysine or glycine. D28 and six of its variants were assayed for bactericidal activity, and all had activity within a twofold dilution of their MIC. One variant (R8) had MICs of 16 µg ml⁻¹ against *E. coli*, 8 µg ml⁻¹ against *B. cereus* and 4 µg ml⁻¹ against *S. aureus* (relative to 64, 16 and 8 µg ml⁻¹, respectively, for D28).

Our linguistic approach to designing synthetic AmPs might be successful because of the pronounced modular nature of natural AmP amino-acid sequences. As we have shown, this approach can be used to expand the AmP sequence space rationally without using structure–activity information or complex simulations of the interactions of a peptide with a membrane. The peptides that we designed are different from previously designed synthetic AmPs¹⁵ in that they bear limited homology to any known protein, which might be desirable for AmPs used in clinical settings. Some researchers argue that widespread clinical use of AmPs that are too similar to human AmPs will inevitably elicit bacterial resistance, compromising our own natural defences and posing a threat to public health¹⁶. In addition, using our approach to develop an arsenal of diverse antimicrobial agents would further reduce concerns about the development of antimicrobial resistance. We hope that this approach will help to expand the diversity of known AmPs well beyond those found in nature, possibly leading to new candidates for AmP-based antibiotic therapeutics.

Our designed AmPs show some degree of homology with natural AmPs because the grammars are based on native sequences. Peptide D28, for example, was matched by grammars derived from 11 natural AmPs, including brevinin, temporin and ponericin. However, Smith-Waterman alignments of our designed peptides against all natural AmPs in the Swiss-Prot/TrEMBL database reveal that the degree of homology is, by design (see Methods), limited. In particular, our two most active peptides, D51 and D28, have only 50 and 60% sequence identity, respectively, with the nearest natural AmP.

Our linguistic design approach might be most valuable as a method for rationally constraining a sequence-based search for new AmPs. Diverse leads generated by our algorithms could be optimized using approaches described in the literature¹⁷. But the

Table 1 | MICs of peptides against bacterial targets

Class	<i>E. coli</i> (gram-negative)		<i>B. cereus</i> (gram-positive)		<i>E. coli</i> or <i>B. cereus</i> MIC
	MIC	MIC	MIC	MIC	
	≤256 µg ml ⁻¹	≤64 µg ml ⁻¹	≤256 µg ml ⁻¹	≤64 µg ml ⁻¹	
Designed	16/40	4/40	8/40	4/40	18/40
Shuffled	1/38	0/38	2/38	1/38	2/38
Natural AmPs	6/8	6/8	4/8	3/8	6/8

Table 2 | MICs of peptides against *S. aureus* and *B. anthracis*

MIC (µg ml ⁻¹)	<i>S. aureus</i>	<i>B. anthracis</i>
8	D28	
16	D51	D28, D51
32		
64	D22	D22
128	D63, S51	D5, D35, D43, D63
256	D5, D35, D43	S51
>256	S5, S22, S28 S35, S43, S63	S5, S22, S28, S35, S43, S63

D, designed on the basis of motifs; S, shuffled control.

linguistic approach described here has a number of limitations. First, sequence families that are poorly conserved on an amino-acid level would not benefit from this approach. Second, we suspect that the small size of AmPs is helpful. Owing to the simple nature of regular grammars, they would be less useful for designing larger proteins and, in particular, proteins with complex tertiary or quaternary structures.

METHODS

Computational design of unnatural AmPs. To design unnatural AmPs, we combinatorially enumerated all grammatical sequences based on the set of ~700 grammars (see Supplementary Materials, Methods, Derivation of grammars). First, for each grammar, we wrote out all possible grammatical amino-acid sequences. So, for example, the grammar [IVL]K[TEGDK]V[GA]K[AELNH][VA][GA]K produced 600 sequences, where $3 \times 5 \times 2 \times 5 \times 2 \times 2 = 600$, owing to the option of choosing one of many amino acids at each bracketed position. There are roughly 3 million such 10-mers that correspond to antimicrobial patterns. Then we wrote out all possible 20-amino-acid sequences for which each window of 10 amino acids is found in the set of 3 million 10-mers. From this set, we removed any 20-mers that had six or more amino acids in a row in common with a naturally occurring AmP. There are roughly 12 million such 20-mers, each of which is a 'tiling' of ten 10-mers.

We clustered these 12 million sequences using the Mcd-hit software¹⁸ at 70% identity. From these clusters, we chose 42 high scoring sequences to test experimentally (see Supplementary Materials, Methods, Scoring of sequences)¹⁹. These sequences have varying degrees of similarity to naturally occurring AmPs, as determined by sequence alignment and discussed in the text.

Peptide synthesis and MIC determination. Fmoc (fluorenylmethoxycarbonyl) chemistry was used to synthesize peptides on the Intavis Multiprep Synthesizer (Intavis LLC) at the MIT Biopolymers Lab, and confirmed using MALDI-TOF Mass Spectrometry and recombinantly produced AmP standards. The MIC was measured with an assay based on the NCCLS M26A and the Hancock assay for cationic peptides¹⁷. See Methods in Supplementary Materials for details on synthesis and antimicrobial assay.

Received 1 May; accepted 4 September 2006.

1. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002).
2. Hancock, R. E. & Patrzykat, A. Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Curr. Drug Targets Infect. Disord.* **2**, 79–83 (2002).

3. Tiozzo, E., Rocco, G., Tossi, A. & Romeo, D. Wide-spectrum antibiotic activity of synthetic, amphipathic peptides. *Biochem. Biophys. Res. Commun.* **249**, 202–206 (1998).
4. Biragyn, A. *et al.* Toll-like receptor 4-dependent activation of dendritic cells by β -defensin 2. *Science* **298**, 1025–1029 (2002).
5. Ellerby, H. M. *et al.* Anti-cancer activity of targeted pro-apoptotic peptides. *Nature Med.* **5**, 1032–1038 (1999).
6. Giangaspero, A., Sandri, L. & Tossi, A. Amphipathic α helical antimicrobial peptides. *Eur. J. Biochem.* **268**, 5589–5600 (2001).
7. Shai, Y. Mode of action of membrane active antimicrobial peptides. *Biopolymers* **66**, 236–248 (2002).
8. Jurafsky, D. & Martin, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Prentice Hall, Upper Saddle River, New Jersey, 2000).
9. Searls, D. B. The language of genes. *Nature* **420**, 211–217 (2002).
10. Hofmann, K., Bucher, P., Falquet, L. & Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219 (1999).
11. Rigoutsos, I. & Floratos, A. Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics* **14**, 55–67 (1998).
12. Wang, Z. & Wang, G. APD: the Antimicrobial Peptide Database. *Nucleic Acids Res.* **32**, D590–D592 (2004).
13. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
14. Wu, M. & Hancock, R. E. W. Interaction of the cyclic antimicrobial cationic peptide bactenecin with the outer and cytoplasmic membrane. *J. Biol. Chem.* **274**, 29–35 (1999).
15. Tossi, A., Sandri, L. & Giangaspero, A. Amphipathic, α -helical antimicrobial peptides. *Biopolymers* **55**, 4–30 (2000).
16. Bell, G. & Gouyon, P.-H. Arming the enemy: the evolution of resistance to self-proteins. *Microbiology* **149**, 1367–1375 (2003).
17. Hilpert, K., Volkmer-Engert, R., Walter, J. & Hancock, R. E. W. High-throughput generation of small antibacterial peptides with improved activity. *Nature Biotechnol.* **23**, 1008–1012 (2005).
18. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein database. *Bioinformatics* **17**, 282–283 (2001).
19. Maizel, J. V. & Lenk, R. P. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl Acad. Sci. USA* **78**, 7665–7669 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank M. Zasloff, K. D. Wittrup, R. Berwick, and G. Georgiou for valuable input on the draft manuscript, and J. Moxley for figure preparation. The authors gratefully acknowledge the support of the Singapore-MIT Alliance, the NIH, and the Fannie and John Hertz Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.S. (gregstep@mit.edu).

LETTERS

Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism

Yuequan Shen¹, Andrzej Joachimiak², Marsha Rich Rosner¹ & Wei-Jen Tang¹

Insulin-degrading enzyme (IDE), a Zn^{2+} -metalloprotease, is involved in the clearance of insulin and amyloid- β (refs 1–3). Loss-of-function mutations of IDE in rodents cause glucose intolerance and cerebral accumulation of amyloid- β , whereas enhanced IDE activity effectively reduces brain amyloid- β (refs 4–7). Here we report structures of human IDE in complex with four substrates (insulin B chain, amyloid- β peptide (1–40), amylin and glucagon). The amino- and carboxy-terminal domains of IDE (IDE-N and IDE-C, respectively) form an enclosed cage just large enough to encapsulate insulin. Extensive contacts between IDE-N and IDE-C keep the degradation chamber of IDE inaccessible to substrates. Repositioning of the IDE domains enables substrate access to the catalytic cavity. IDE uses size and charge distribution of the substrate-binding cavity selectively to entrap structurally diverse polypeptides. The enclosed substrate undergoes conformational changes to form β -sheets with two discrete regions of IDE for its degradation. Consistent with this model, mutations disrupting the contacts between IDE-N and IDE-C increase IDE catalytic activity 40-fold. The molecular basis for substrate recognition and allosteric regulation of IDE could aid in designing IDE-based therapies to control cerebral amyloid- β and blood sugar concentrations^{1,8,9}.

Insulin-degrading enzyme (IDE), originally identified by its ability to rapidly degrade insulin, is a highly conserved Zn^{2+} -metalloprotease found in bacteria, fungi, plants and animals¹⁰ (Supplementary Fig. 1). In budding yeast, IDE homologues have key roles in bud-site selection and mating^{11,12}. IDE is unusual in its high affinity for substrates that are highly diverse in sequence and structure¹ (Supplementary Fig. 2). Furthermore, IDE has a remarkable capacity to cleave selectively some hormones without degrading related family members^{1,2} (Supplementary Fig. 2). Despite this high affinity, IDE cleaves its substrates several times and the recognition motifs are not obvious from the cleavage sites. The molecular basis by which IDE shows high selectivity but degenerate cleavage sites for a broad range of hormones has so far remained elusive.

To understand the substrate recognition and catalytic mechanism of IDE, we solved crystal structures of the 113-kDa human, Zn^{2+} -bound, catalytically inactive IDE mutant IDE-E111Q (ref. 13) in complex with insulin B chain at 2.25 Å resolution (Fig. 1a), and Zn^{2+} -free IDE-E111Q in complex with amyloid- β peptide ($\text{A}\beta$ (1–40)), amylin and glucagon at 2.1 Å, 2.6 Å and 2.5 Å resolution, respectively (Fig. 2). Each IDE monomer comprises four structurally homologous $\alpha\beta$ roll domains (domain 1, residues 43–285; domain 2, residues 286–515; domain 3, residues 542–768; and domain 4, residues 769–1,016), which share less than 25% sequence similarity (Fig. 1b and Supplementary Figs 3 and 4). The N-terminal domains 1 and 2 form a $\alpha\beta\alpha\beta$ sandwich (IDE-N), as do domains 3 and 4 (IDE-C)¹⁴.

IDE-N and IDE-C are joined by a 28-residue extended loop and form an enclosed chamber, shaped like a triangular prism, with triangular base dimensions of $35 \times 34 \times 30$ Å and a height of 36 Å. This enclosed cavity has a total volume of $\sim 1.3 \times 10^4$ Å³, which is just large enough to encapsulate insulin A and B chains (Fig. 1c and Supplementary Fig. 5). All four domains contribute surface to the internal chamber. The surface provided by IDE-N is largely neutral or negatively charged, whereas that provided by IDE-C is predominantly positively charged (Fig. 1d). IDE domain 1 contains the catalytic site with a Zn^{2+} ion coordinated by two histidines (residues 108 and 112) and one glutamate (residue 189; ref. 15 and Fig. 1a, e). In our IDE structures, the general base glutamate 111 is replaced with glutamine to prevent the catalysis of substrates¹³.

Two discrete segments of four sequence diverse substrates—insulin B chain, the amyloid- β peptide ($\text{A}\beta$ (1–40)), amylin and glucagon—are clearly visible in structures of the IDE–substrate complex, and they share similar features (Fig. 2 and Supplementary Fig. 6). The N-terminal 3–5 amino acids and cleavage-site-containing 7–13 amino acids of all four substrates form β -sheets with IDE strands β 12 and β 6, respectively (Fig. 2). The remaining regions (55–72%) of all four substrates in our IDE–substrate complexes are disordered, even though they are present in the chamber, as verified by mass spectrometry analysis (Supplementary Fig. 7). At the catalytic site, several residues of IDE domains 1 and 4 form a largely polar cavity with patches of hydrophobic and charged regions that interact with cleavage sites in all four substrates (Fig. 2 and Supplementary Fig. 8). The bulky hydrophobic residues at the P1 residues of the IDE substrates interact with Phe 141 at the S1 subsite of IDE domain 1, whereas the hydrophobic residues of the P1' residues are buried deeply in the hydrophobic patch surrounding the S1' subsite of IDE domain 1. In addition, Arg 824 and Tyr 831 of IDE domain 4 form hydrogen bonds with the P1 and P1' residues of substrates. Mutations of these two residues to alanine substantially reduce the catalytic rate of IDE (Fig. 3a and Supplementary Fig. 9). This is consistent with the idea that IDE-N functions as the catalytic domain, whereas IDE-C facilitates the binding of substrates¹⁴.

The enclosed substrate-binding compartment of IDE prevents the entry and exit of substrates. Thus, IDE needs to undergo a significant conformational change from the open state, which can accept substrate, to the closed state for proper substrate recognition and catalysis. The structural comparison of substrate-bound IDE with substrate-free *Escherichia coli* pitrilysin (PDB accession code 1Q2L) reveals how repositioning between IDE-N and IDE-C can lead to the open state, which allows substrate to access the catalytic cavity¹⁶ (Fig. 3b and Supplementary Fig. 10). Pitrilysin shares 25% sequence identity with IDE and is arranged as two globular entities, pitrilysin-N and pitrilysin-C (Fig. 3b). Structural comparison of pitrilysin and IDE reveals that pitrilysin-C rotates 54° away from pitrilysin-N such

¹Ben-May Institute for Cancer Research, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. ²Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.

that domain 4 of pitrilysin does not contact domain 1. Thus, IDE may normally equilibrate between the substrate-free open state (IDE^o) and the closed state (IDE^c; Fig. 3c). IDE^c cannot bind substrates, but it is capable of degrading substrate after substrates are entrapped inside the catalytic chamber. IDE^o can bind substrates; however, without the contribution of residues from IDE-C (Arg 824 and Tyr 831 of domain 4), IDE^o is less active than IDE^c (ref. 14). After the transition from IDE^o to IDE^c, the entrapped peptides need to fit into the catalytic cleft of IDE to enable their cleavage once or multiple times before the reopening of IDE.

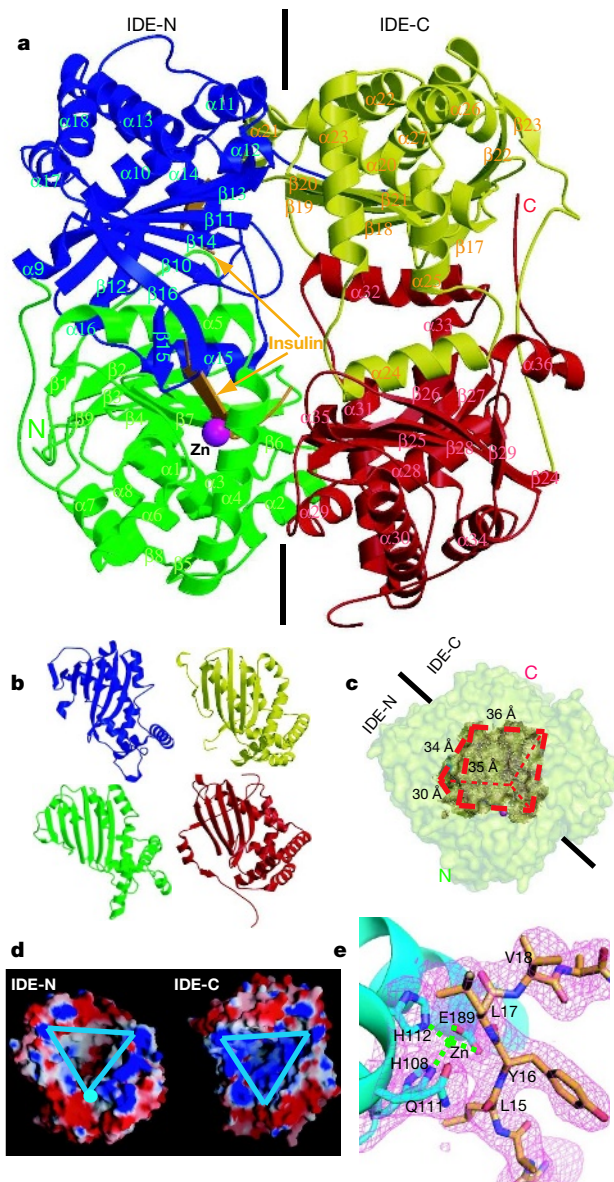


Figure 1 | Overall structure of IDE-E111Q in complex with insulin B chain. **a**, Secondary structure representation of the IDE-E111Q-insulin B chain complex. Domains 1, 2, 3 and 4 are coloured green, blue, yellow and red, respectively. Zn²⁺ and insulin B chain are coloured magenta and orange, respectively. **b**, Structure homology of the four domains of IDE. **c**, Surface representation of the substrate-binding chamber of IDE. The outer surface of IDE and the substrate chamber are coloured pale yellow and brown, respectively. **d**, Electrostatic surface representation of the IDE substrate-binding chamber. The inner substrate binding chambers of IDE-N and IDE-C are marked by triangles. The surface is coloured as follows: negative, red; positive, blue; neutral, white. **e**, Catalytic centre of IDE. The simulated annealing omit map (magenta) is contoured at 3.5σ. IDE and insulin B chain are coloured cyan and orange, respectively.

IDE-N and IDE-C have extensive interactions that bury a large surface (11,496 Å²) with good shape complementarity¹⁷ (shape complementarity score=0.66) and numerous hydrogen bonds (Supplementary Fig. 4e). This leads to the prediction that, in the absence of interactions with other proteins or factors, the substrate-free IDE^c state is stable and the catalytic chamber of IDE is mostly closed. To test this hypothesis, we constructed three IDE mutants carrying double cysteine mutations that could loosen the contacts between IDE-N and IDE-C, therefore promoting opening of the catalytic chamber and increasing the catalytic rate (Fig. 3d). The two cysteine mutations are also designed to form a potential disulphide bond; thus, these mutants could be locked in the closed conformation¹⁸. When fluorogenic substrate V was used as a substrate, all three IDE mutants—D426C/K899C, N184C/Q828C and S132C/E817C—had 30–40-fold higher catalytic activity than did wild-type IDE in the presence of the reducing agent Tris-(2-carboxyethyl)-phosphine (TCEP; Fig. 3d). The increased proteolytic activities of these three IDE mutants were confirmed when either insulin or the amyloid-β peptide (1–42) were used (Supplementary Fig. 11). We then tested whether these IDE mutants could be inactivated in the presence of the oxidizing agent K₃Fe(CN)₆, which facilitates disulphide bond formation. We found that S132C/E817C and N184C/Q828C were mostly inactive in the presence of K₃Fe(CN)₆ (Fig. 3e, f, and Supplementary Fig. 9). The activities of these two mutants were restored when TCEP was added (Fig. 3f). The third pair, D426C/K899C, was not inactivated by K₃Fe(CN)₆, presumably owing to the lack of the formation of a disulphide bond. As a control, wild-type IDE was insensitive to the treatment of TCEP or K₃Fe(CN)₆ (Fig. 3d and Supplementary Fig. 9). The oligomeric state of these three IDE mutants is similar to that of wild-type IDE, excluding the idea that a difference in oligomerization is the cause of their hyperactivity¹⁹ (Supplementary Fig. 12).

The comparison of IDE-free and IDE-bound insulin B chain, Aβ(1–40), amylin and glucagon reveals substantial conformational changes in IDE substrates on binding to IDE^{20–25} (Fig. 4a). Both the N-terminal loop and the α-helical cleavage site turn into β-strands. IDE cleaves insulin B chain and amyloid-β at several sites^{1,26}. The binding of insulin B chain and Aβ(1–40) to the IDE catalytic cleft positions both substrates for cleavage at known sites (Fig. 4b). IDE also cleaves amylin and glucagon at several locations, but these sites had not previously been identified. Our matrix-assisted laser desorption–ionization time-of-flight analysis verified that the cleavage sites for amylin and glucagon do indeed correspond to degradation sites shown by the crystal structures of the IDE–substrate complex (Fig. 4b and Supplementary Fig. 13).

Together, our structural and biochemical analyses reveal that at least four factors contribute to the unique mechanism of substrate recognition by IDE. Favourable binding of the substrate N terminus and cleavage sites to β-strands within IDE and proper anchoring of the cleavage site in the catalytic cleft are clearly key specific determinants. In addition, peptides that do not have significant positive charges at the C terminus and avoid the charge repulsion from IDE-C are better IDE substrates. Our IDE–substrate structures show that the C termini of insulin B chain, amyloid-β and amylin make substantial contacts with the IDE inner cavity, which is highly positively charged (Fig. 1d). Brain natriuretic peptide, glucagon-like peptide and insulin growth factor I (IGF-I), which have many positively charged residues at their C termini, are poor substrates (Supplementary Fig. 2). By contrast, the related hormones atrial natriuretic peptide, glucagon and IGF-II, which lack positive charges at their C termini, are excellent IDE substrates. The fourth determining factor is size. The catalytic chamber of IDE is large enough to accommodate only relatively small peptides (estimated to be <50 amino acids). Larger peptides such as transforming growth factor-β (TGF-β) and pro-insulin are less likely to be entrapped by IDE than the related, smaller hormones, TGF-α and insulin. Consequently, the degradation of such larger peptides is considerably slower¹ (Supplementary Fig. 2).

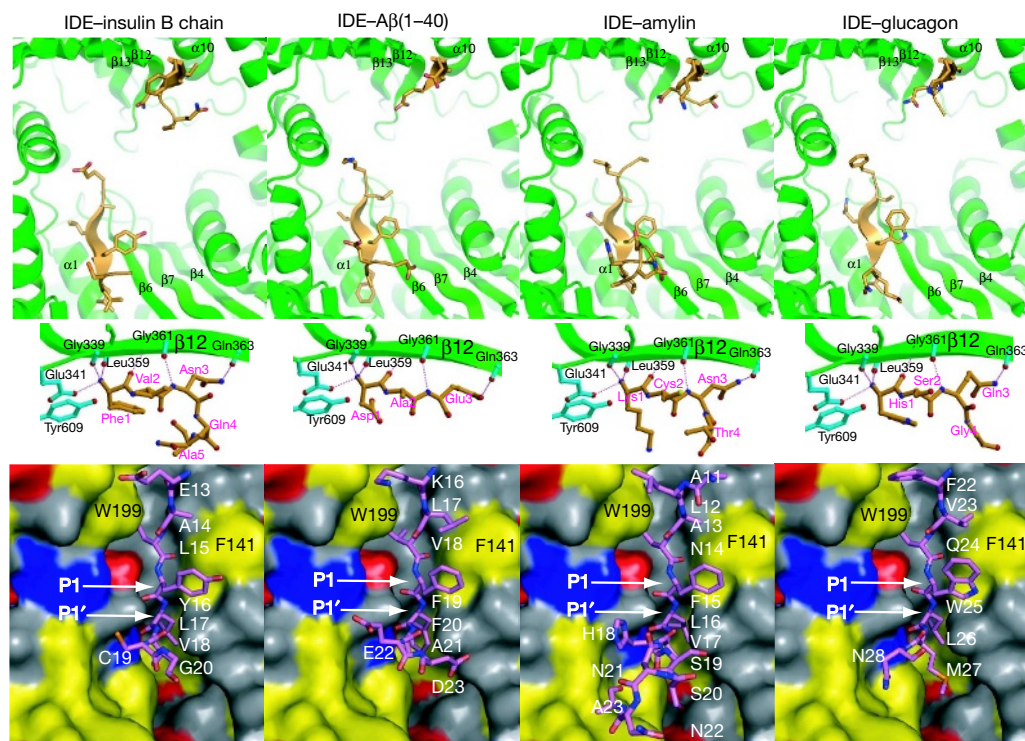


Figure 2 | Interaction of IDE with its substrates. Top, global view of how substrates bind to IDE. Middle, detailed interaction of the N terminus of substrates with IDE. Secondary structures and substrates of IDE are coloured green and orange, respectively. Bottom, interaction of the cleavage

site of IDE substrates with the IDE catalytic cleft. The polar, hydrophobic, positive and negative surfaces are coloured grey, yellow, blue and red, respectively.

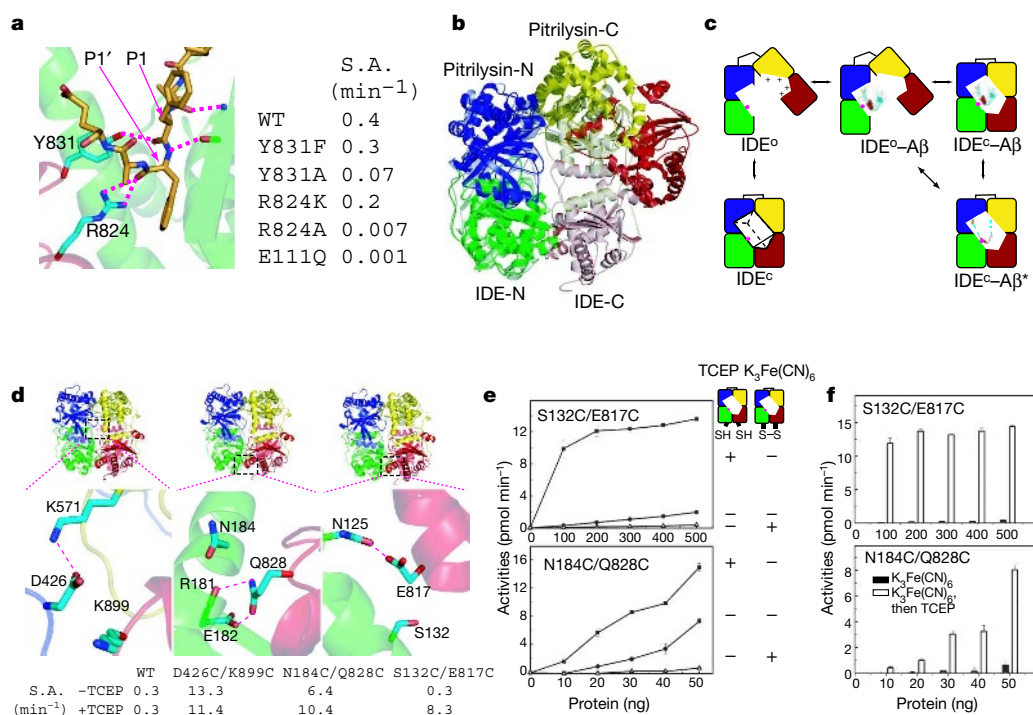


Figure 3 | Conformational switch of IDE. **a**, Details of the interaction with Aβ(1-40), showing residues Tyr 831 and Arg 824 from domain 4 of IDE and the specific activity (S.A.) of IDE mutants. Domains 1 and 4 of IDE are coloured green and red, respectively, and Aβ(1-40) is coloured orange. **b**, Structural comparison of IDE with pitrilysin. Pitrilysin is shown in solid colours and IDE in transparent colours. The colour scheme is the same as Fig. 1a. Domains 1 and 2 of IDE and pitrilysin are superimposed. **c**, Model of

how IDE binds to its substrate. IDE has two conformational states: the open state, IDE^o; and the closed state, IDE^c. **d**, Details of the interactions of residues between IDE-N and IDE-C at three discrete locations and the catalytic activities of three IDE double cysteine mutants in the presence or absence of TCEP. **e**, Activities of two double cysteine mutants in the presence of TCEP or K₃Fe(CN)₆. **f**, Rescue of S132C/E817C and N184C/Q828C by TCEP. Results in **e** and **f** are the mean ± s.e.m.

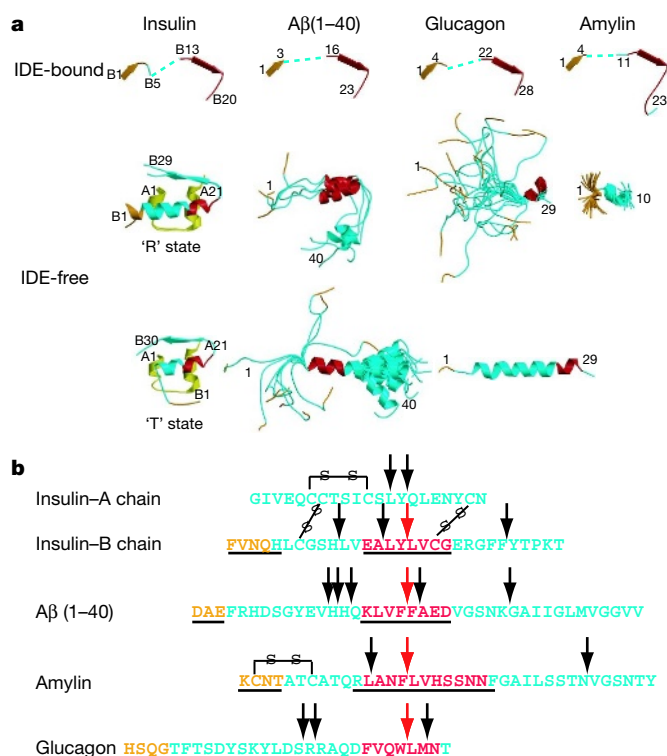


Figure 4 | Conformational changes and catalysis of IDE substrates.

a, Secondary structure of IDE substrates in the IDE-bound (top) or free (bottom) form. The N terminus and IDE catalytic cleft binding segment are coloured orange and red, respectively. The PDB accession codes for insulin are 1G7A and 1ZEH, those for Aβ are 1AML and 1BA4, those for glucagon are 1GCN and 1KX6, and that for amylin is 1KUW. **b**, Sequence comparison of four IDE substrates. Arrows indicate the main cleavage sites of the substrate by IDE^{1,26} (Supplementary Fig. 13). Amino acids that are underlined are observed in the crystal structures of substrate-bound IDE.

IDE has a unique regulatory mechanism. As an M16A member of the Zn²⁺-metalloprotease family, it shares similar secondary structure and domain organization with yeast mitochondria processing peptidase (MPP), a distally related M16B member²⁷ (Supplementary Fig. 10). Similar to IDE, MPP also uses the exosite for substrate recognition^{16,27,28}. The catalytic chamber of MPP stays open, however, whereas IDE has a buried catalytic site in the structure and access to this chamber is kinetically controlled by the closed–open conformational switch. IDE can also self-oligomerize and the interaction between two IDE dimers could lock IDE in the IDE^c state (Supplementary Fig. 14). This might explain how oligomerization allosterically regulates the catalytic activity of IDE¹⁹. Small molecules that could shift the equilibrium between IDE^c and IDE^o towards the open state or reduce IDE oligomerization will probably allosterically regulate the activity of IDE. Such compounds might facilitate the clearance of amyloid-β and other pathologically relevant IDE substrates.

METHODS

Protein preparation and crystallization. We purchased human insulin (RayBiotech), Aβ(1–40) (Biosource), amylin (Bachem) and glucagon (Anaspec). Human IDE-E111Q and selenomethionyl-IDE-E111Q were expressed in *E. coli* strains Rosetta(DE3) and B834(DE3)pUBS520, respectively, at 25 °C by IPTG induction for 19 h, and purified by Ni²⁺-NTA, source-Q and Superdex S-200 columns. Preformed IDE–substrate complexes isolated from S-200 columns (~15 mg ml⁻¹ in 20 mM Tris-HCl and 50 mM NaCl; pH 8.0) in the presence of a reducing agent, 1 mM TCEP, were mixed with equal volumes of reservoir solution containing 0.1 M HEPES (pH 7.0), 12% (w/v) PEGMME-5000, 5% tacsimate and 10% dioxane. Crystals appeared after 1–3 weeks at 18 °C and were then equilibrated in cryo-protective buffer containing well buffer and 30% glycerol. IDE–substrate complex crystals belong to the space group P6₃, with

unit-cell dimensions $a=b=262$ Å and $c=90$ Å, and contain a dimeric IDE–substrate complex per asymmetric unit.

Structure determination. Data were collected at 14-BM-C and 19-ID stations in the Advanced Photon Source at Argonne National Laboratory and processed with HKL2000. Anomalous diffraction data were collected on crystals of the selenomethionyl-IDE–insulin-B-chain complex, and 34 out of 52 selenium sites were located by the Shake-and-Bake program. Initial phases were obtained by SAD using SHARP²⁹. DM programs and phase extension were done on the Zn²⁺-bound IDE–insulin-B-chain complex (Supplementary Table 1 and Fig. 15). AMoRe was used to obtain the initial phases of structures of Zn²⁺-free IDE in complex with insulin B chain, Aβ, amylin and glucagon by using the IDE–insulin B chain complex as a template. Model building and refinement of IDE–substrate complexes were done with COOT and CNS³⁰. The final structures of Zn²⁺-IDE–insulin-B-chain, Zn²⁺-free IDE–insulin-B-chain, IDE–Aβ, IDE–amylin and IDE–glucagon have an R_{free} value of 23.3%, 22.5%, 22.3%, 22.5% and 22.5%, and an R_{cryst} value of 20.6%, 20.5%, 20.3%, 19.6% and 19.8%, respectively. Detailed data collection and refinement statistics are summarized in Supplementary Table 2. The electron density of the whole IDE dimer (residues 43–1,016) is clearly visible, except for a short disordered loop (residues 974–976) and the C-terminal end (residues 1,017–1,018). Only the structure of the Zn²⁺-bound IDE–insulin-B-chain complex is discussed because the structure of the Zn²⁺-free IDE–insulin-B-chain complex has less clear electron density for the side chains of insulin B chain at the catalytic cleft (Supplementary Fig. 6) and crystals of the complex between IDE and intact insulin did not diffract well (Supplementary Fig. 16).

Enzymatic assay. IDE mutants were constructed with a Quik-Change kit (BioLabs Manufacturing) and purified by Ni²⁺-NTA and source-Q columns. Enzyme activities were assayed by mixing 100 μl of 5 μM fluorogenic peptide substrate V (R&D Systems) in 50 mM potassium phosphate (pH 7.3) and 5 μl of IDE proteins at 37 °C at the indicated time and by monitoring fluorescence intensity on a Tecan Safire2 microplate reader at an excitation wavelength of 327 nm and an emission wavelength of 395 nm (ref. 14). To measure the activities of two double cysteine mutants in the presence of TCEP or K₃Fe(CN)₆, the indicated quantities of S132C/E817C and N184C/Q828C were pre-incubated with 1 mM TCEP or 1 mM K₃Fe(CN)₆ at room temperature for 10 and 60 min, respectively, to carry out reducing or oxidizing reactions. The fluorogenic substrate was then added and incubated at 37 °C for 30 min. To perform the rescue experiment of S132C/E817C and N184C/Q828C by TCEP, the activities of these two mutants in the presence of 1 mM K₃Fe(CN)₆ were first measured after a 30-min incubation. TCEP was then added to a concentration of 5 mM and the activities were measured after 30 or 60 min for S132C/E817C or N184C/Q828C, respectively.

Received 14 March 2006; accepted 7 August 2006.

Published online 11 October 2006.

- Duckworth, W. C., Bennett, R. G. & Hamel, F. G. Insulin degradation: progress and potential. *Endocr. Rev.* **19**, 608–624 (1998).
- Kurochkin, I. V. Insulin-degrading enzyme: embarking on amyloid destruction. *Trends Biochem. Sci.* **26**, 421–425 (2001).
- Bennett, R. G., Duckworth, W. C. & Hamel, F. G. Degradation of amylin by insulin-degrading enzyme. *J. Biol. Chem.* **275**, 36621–36625 (2000).
- Farris, W. *et al.* Insulin-degrading enzyme regulates the levels of insulin, amyloid β-protein, and the β-amyloid precursor protein intracellular domain *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 4162–4167 (2003).
- Farris, W. *et al.* Partial loss-of-function mutations in insulin-degrading enzyme that induce diabetes also impair degradation of amyloid β-protein. *Am. J. Pathol.* **164**, 1425–1434 (2004).
- Leissring, M. A. *et al.* Enhanced proteolysis of β-amyloid in APP transgenic mice prevents plaque formation, secondary pathology, and premature death. *Neuron* **40**, 1087–1093 (2003).
- Miller, B. C. *et al.* Amyloid-β peptide levels in brain are inversely correlated with insulin activity levels *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 6221–6226 (2003).
- Selkoe, D. J. Clearing the brain's amyloid cobwebs. *Neuron* **32**, 177–180 (2001).
- Tanzi, R. E. & Bertram, L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* **120**, 545–555 (2005).
- Mirsky, I. A. & Broth-Kahn, R. H. The inactivation of insulin by tissue extracts. I. The distribution and properties of insulin inactivating extracts (insulinase). *Arch. Biochem.* **20**, 1–9 (1949).
- Adames, N., Blundell, K., Ashby, M. N. & Boone, C. Role of yeast insulin-degrading enzyme homologs in propheromone processing and bud site selection. *Science* **270**, 464–467 (1995).
- Fujita, A. *et al.* A yeast gene necessary for bud-site selection encodes a protein similar to insulin-degrading enzymes. *Nature* **372**, 567–570 (1994).
- Perlman, R. K. & Rosner, M. R. Identification of zinc ligands of the insulin-degrading enzyme. *J. Biol. Chem.* **269**, 33140–33145 (1994).

14. Li, P., Kuo, W.-L., Yousef, M., Rosner, M. R. & Tang, W.-J. The C-terminal domain of human insulin degrading enzyme is required for the dimerization and substrate recognition. *Biochem. Biophys. Res. Commun.* **343**, 1032–1037 (2006).
15. Becker, A. B. & Roth, R. A. An unusual active site identified in a family of zinc metalloendopeptidases. *Proc. Natl Acad. Sci. USA* **89**, 3835–3839 (1992).
16. Maskos, K. in *Handbook of Metalloproteins* (eds Messerschmidt, A., Dode, W. & Cygler, M.) 190–198 (John Wiley & Sons, 2004).
17. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
18. Dombkowski, A. A. Disulfide by Design: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics* **19**, 1852–1853 (2003).
19. Song, E. S., Juliano, M. A., Juliano, L. & Hersh, L. B. Substrate activation of insulin-degrading enzyme (insulysin). A potential target for drug development. *J. Biol. Chem.* **278**, 49789–49794 (2003).
20. Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. Transmission of conformational change in insulin. *Nature* **302**, 500–505 (1983).
21. Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. X-ray analysis of glucagon and its relationship to receptor binding. *Nature* **257**, 751–757 (1975).
22. Braun, W., Wider, G., Lee, K. H. & Wuthrich, K. Conformation of glucagon in a lipid–water interphase by ^1H nuclear magnetic resonance. *J. Mol. Biol.* **169**, 921–948 (1983).
23. Coles, M., Bicknell, W., Watson, A. A., Fairlie, D. P. & Craik, D. J. Solution structure of amyloid β -peptide(1–40) in a water-micelle environment. Is the membrane-spanning domain where we think it is? *Biochemistry* **37**, 11064–11077 (1998).
24. Mascioni, A., Porcelli, F., Ilangoan, U., Ramamoorthy, A. & Veglia, G. Conformational preferences of the amylin nucleation site in SDS micelles: an NMR study. *Biopolymers* **69**, 29–41 (2003).
25. Sticht, H. *et al.* Structure of amyloid A4-(1–40)-peptide of Alzheimer's disease. *Eur. J. Biochem.* **233**, 293–298 (1995).
26. Leissring, M. A. *et al.* Kinetics of amyloid β -protein degradation determined by novel fluorescence- and fluorescence polarization-based assays. *J. Biol. Chem.* **278**, 37314–37320 (2003).
27. Taylor, A. B. *et al.* Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure* **9**, 615–625 (2001).
28. Luciano, P. & Geli, V. The mitochondrial processing peptidase: function and specificity. *Experientia* **52**, 1077–1082 (1996).
29. La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
30. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of APS SBC 19-ID and Biocars 14-BM-C, P. Li and Q. Guo for help with data collection; M. Manolopoulou for help with protein purification; H. Im for help with IDE enzymatic assays; D. Wolfgeher for help with mass spectrometry; E. Johnson for providing the synthetic A β (1–42) peptide; and R. Bourdeau for critically reading the manuscript. This research was supported by NIH and The University of Chicago Diabetes Center grants to W.T.

Author Contributions W.T. designed the experiments and Y.S. conducted all experiments. W.T. and Y.S. analysed the results and co-wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Coordinates of the X-ray structures of the substrate-bound IDE have been deposited in the RCSB Protein Data Bank under accession code 2G54 (Zn $^{2+}$ -IDE-insulin-B-chain), 2G56 (Zn $^{2+}$ -free IDE-insulin-B-chain), 2G47 (IDE-A β (1–40)), 2G48 (IDE-amylin) and 2G49 (IDE-glucagon). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.T. (wtang@uchicago.edu).

Direct observation of individual RecA filaments assembling on single DNA molecules

Roberto Galletto^{1,2,3}, Ichiro Amitani^{1,2,3}, Ronald J. Baskin² & Stephen C. Kowalczykowski^{1,2,3}

Escherichia coli RecA is essential for the repair of DNA double-strand breaks by homologous recombination¹. Repair requires the formation of a RecA nucleoprotein filament. Previous studies have indicated a mechanism of filament assembly whereby slow nucleation of RecA protein on DNA is followed by rapid growth^{2–7}. However, many aspects of this process remain unclear, including the rates of nucleation and growth and the involvement of ATP hydrolysis, largely because visualization at the single-filament level is lacking. Here we report the direct observation of filament assembly on individual double-stranded DNA molecules using fluorescently modified RecA. The nucleoprotein filaments saturate the DNA and extend it 1.6-fold. At early time points, discrete RecA clusters are seen, permitting analysis of single-filament growth from individual nuclei. Formation of nascent RecA filaments is independent of ATP hydrolysis but is dependent on the type of nucleotide cofactor and the RecA concentration, suggesting that nucleation involves binding of 4–5 ATP–RecA monomers to DNA. Individual RecA filaments grow at rates of 3–10 nm s^{−1}. Growth is bidirectional and, in contrast to nucleation, independent of nucleotide cofactor, suggesting addition of 2–7 monomers s^{−1}. These results are in accord with extensive genetic and biochemical studies, and indicate that assembly *in vivo* is controlled at the nucleation step. We anticipate that our approach and conclusions can be extended to the related eukaryotic counterpart, Rad51 (ref. 8), and to regulation by assembly mediators^{9–11}.

Recombinational DNA repair involves the formation of a RecA nucleoprotein filament typically on single-stranded DNA (ssDNA) for pairing and exchange with homologous double-stranded DNA (dsDNA)⁸. Although the kinetically preferred substrate is ssDNA, RecA can form filaments on dsDNA^{2,3} that are active in ‘inverse’ DNA strand exchange with complementary ssDNA^{12,13}. Previous studies indicated that filament assembly involves nucleation of RecA onto DNA followed by elongation (Supplementary Fig. 1), but direct visualization of the time-dependent evolution of an individual RecA nucleoprotein filament has been lacking. To visualize RecA filament formation directly, we prepared a fluorescently modified RecA protein that retains all biochemical activity (Supplementary Figs 2–4). This protein was used to monitor RecA filament formation on dsDNA using the following strategy¹⁴. A flow cell is used to generate three separate laminar flow channels (Fig. 1a). Biotinylated bacteriophage λ DNA, attached to streptavidin-coated beads (1 μ m), is optically trapped in the capture channel. To ensure that the single DNA molecule is at full length, trapping is performed using the fluorescent DNA-binding dye YOYO-1. The trapped bead–YOYO-1–DNA complex is then moved into a second channel (the observation channel) under conditions that dissociate the dye from DNA. Figure 1b (panels 1 and 2) shows a λ DNA molecule in the observation channel before and after dissociation of YOYO-1, respectively. This dye-free DNA is subsequently moved to a third channel (RecA channel) containing the fluorescently modified RecA. After incubation to allow for the formation of a RecA filament,

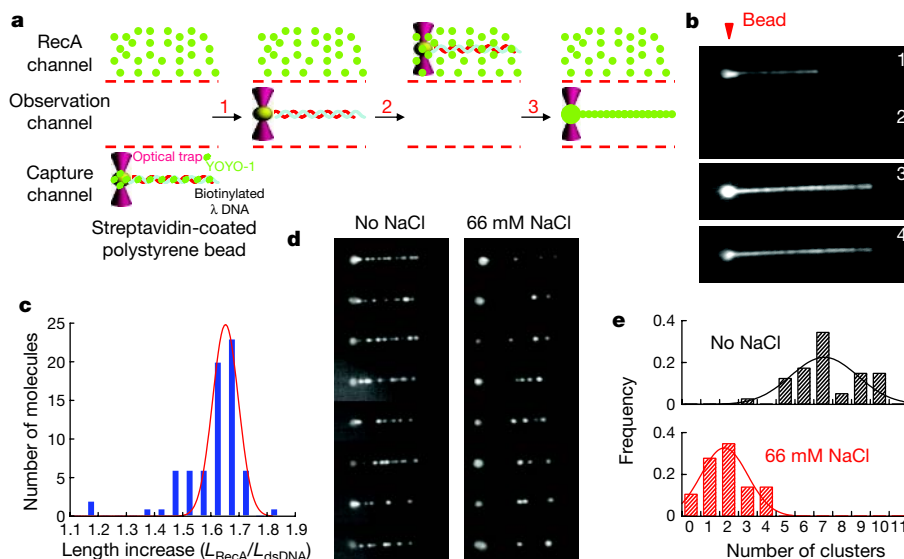


Figure 1 | Nucleation and growth of RecA filaments visualized on individual dsDNA molecules. **a**, Experimental approach (see the text for details). **b**, Panels 1 and 2 show a bead–DNA complex trapped before (panel 1) and after (panel 2) dissociation of YOYO-1. Panel 3 shows a fluorescent filament in the observation channel after 5 min in the RecA channel with 500 nM RecA, 0.5 mM ATP- γ S and 1 mM Mg(OAc)₂. Panel 4 shows the same RecA filament after ~20 min in the observation channel with continued flow, no illumination. **c**, Relative extension for RecA filaments ($n = 72$); the line is a gaussian fit. **d**, Eight different nascent RecA filaments after incubation for 1 min with 200 nM RecA, 0.5 mM ATP- γ S, 1 mM Mg(OAc)₂ and indicated NaCl concentration. **e**, Number of clusters observed; lines are gaussian fits.

¹Section of Microbiology and ²Section of Molecular & Cellular Biology, ³Center for Genetics and Development, University of California, Davis, California 95616, USA.

the trapped DNA is moved back into the observation channel where the bound fluorescent RecA is visualized directly.

After incubation in the RecA channel for 5 min in the presence of ATP- γ S, a complete RecA nucleoprotein filament is visible in the observation channel (Fig. 1b, panel 3). The nucleoprotein filament is extended relative to naked dsDNA, consistent with the extension induced by RecA¹⁵. The size distribution for the relative extension of DNA by RecA (Fig. 1c) is approximately gaussian, with an average extension of 1.65 ± 0.05 ($n = 72$). Given that the persistence lengths of dsDNA and RecA nucleoprotein filaments differ greatly^{4,5,15}, the observed extension is consistent with the polymorphic ~ 1.5 -fold increase in dsDNA length reported, and indicates that complete RecA filaments are formed. The length of the RecA nucleoprotein filament does not change after 20 min of continued flow in the RecA- and ATP- γ S-free observation channel (Fig. 1b, panel 4), even in the presence of NaCl or ATP (Supplementary Fig. 5), indicating that dissociation of the ATP- γ S-RecA-dsDNA complex is negligible.

Although our protocol does not permit the monitoring of RecA filament formation in real time due to the high background fluorescence in the RecA channel, snapshots of filament formation on the same DNA molecule can be obtained by incubating the DNA in the RecA channel for a short time (dipping), and then moving the DNA to the observation channel. Examples of eight partial filaments formed by dipping in the RecA channel for 1 min are shown in Fig. 1d. It is evident that multiple RecA clusters have formed on the dsDNA, showing directly that RecA filaments form from multiple nucleation events. The frequency of RecA nucleation is strongly dependent on salt concentration (Fig. 1e). From 20 to 80 mM NaCl, the average rate of RecA cluster formation decreased from ~ 8 to $\sim 1 \text{ min}^{-1}$ (Supplementary Fig. 6); given that λ DNA is 48,502 base pairs (bp), the average nucleation frequency ranged from $\sim 1.6 \times 10^{-4}$ to $2 \times 10^{-5} \text{ nuclei min}^{-1} \text{ bp}^{-1}$. Previous measurements, at different conditions, produced comparable values^{4,16,17}, but, as shown below, nucleation frequency is strongly influenced by reaction conditions. The affinity of RecA for DNA is highly sensitive to salt concentration^{18,19}, suggesting that nucleation is occurring on the DNA. This conclusion is in agreement with a previous ensemble study⁷ and is consistent with the DNA composition-dependence and nucleotide-dependence presented below. Furthermore, even though nucleation was seemingly equally likely along the length of the dsDNA at the lowest NaCl concentration (Supplementary Fig. 7), nucleation occurred more frequently and preferentially at (A+T)-rich regions of the λ DNA when the NaCl concentration was elevated (Supplementary Figs 8–10). Ensemble studies showed that (A+T)-rich duplex DNA stimulates nucleoprotein filament assembly, which was attributed to a rate-limiting nucleation of RecA onto one strand of the dsDNA².

The affinity of RecA for DNA is strongly influenced by the nature of the bound nucleotide cofactor^{18,19}. Although ATP- γ S-RecA clusters are stable for long times in the observation channel, filaments formed with ATP completely dissociated in ~ 6 min after a lag of ~ 10 – 20 s (our own unpublished observations). However, if the ATP-RecA clusters are moved into an observation channel containing ATP- γ S, then they are completely stabilized (Fig. 2a). The average number of clusters formed with ATP- γ S or ATP, under otherwise identical experimental conditions, is shown in Fig. 2b. With ATP- γ S, RecA nucleates ~ 5 -fold faster than with ATP ($10.2 \pm 0.6 \text{ min}^{-1}$ compared with $1.86 \pm 0.06 \text{ min}^{-1}$), consistent with previous modelling⁴. The difference in nucleation rate cannot be attributed to a different degree of saturation of RecA with ATP or ATP- γ S due to differences in affinity, because identical rates are obtained when a fourfold higher concentration of ATP is used (Fig. 2b, orange diamonds). However, the difference might be the consequence of ATP-hydrolysis-induced dissociation. To address this point, Ca^{2+} was substituted for Mg^{2+} in the RecA channel, because Ca^{2+} reduces the rate of ATP hydrolysis by 96%¹⁸. Although Ca^{2+} decreases the rate of nucleation, the same ~ 5 -fold difference in rate is maintained

when ATP- γ S replaces ATP (Fig. 2b). Thus, the nucleation frequency is not sensitive to ATP hydrolysis but rather is sensitive to the differences in affinity for DNA that are induced by nucleotide cofactors and metal ions^{18,19}. This conclusion, together with the increased nucleation frequency in (A+T)-rich regions (Supplementary Figs 7–10), further implies that nucleation occurs on the DNA.

The effect of RecA concentration on nucleus formation is shown in Fig. 2c. The rate of cluster formation, determined from the linear increase in nucleation frequency with time, is shown as a function of RecA concentration in Fig. 2d. A pronounced nonlinear dependence on RecA concentration is evident. The solid line is the fit according to the relationship $k_{\text{obs}} = c[\text{RecA}]^n$, where n represents the minimum number of RecA molecules involved in the cooperative process and c is a constant²⁰. This simplistic approach suggests that formation of a stable nucleus on dsDNA requires at least four RecA molecules ($n = 4 \pm 0.4$). Notably, the same dependence ($n = 4.5 \pm 0.5$) is observed with ATP (Fig. 2d). Although the same power dependence is maintained, the entire curve is shifted towards higher RecA concentrations, consistent with the slower nucleation rate of ATP-RecA. This simple analysis suggests that 4–5 RecA molecules are needed to form a stable nucleus.

At high concentrations of salt, the nucleation frequency is strongly limited and growth from an individual cluster can be monitored via successive incubations in the RecA channel (Figs 3 and 4). An individual DNA molecule was dipped for 1 min in the ATP- γ S-RecA channel until clusters were detected (Fig. 3a, 0 min). Then, the molecule was repeatedly moved into the RecA channel for a fixed time, and back out to the observation channel to follow growth of the initial clusters (Fig. 3a, 1–9 min). The size of the two clusters in the top panel (0 min) increases linearly with observed rates of 3.3 ± 0.1

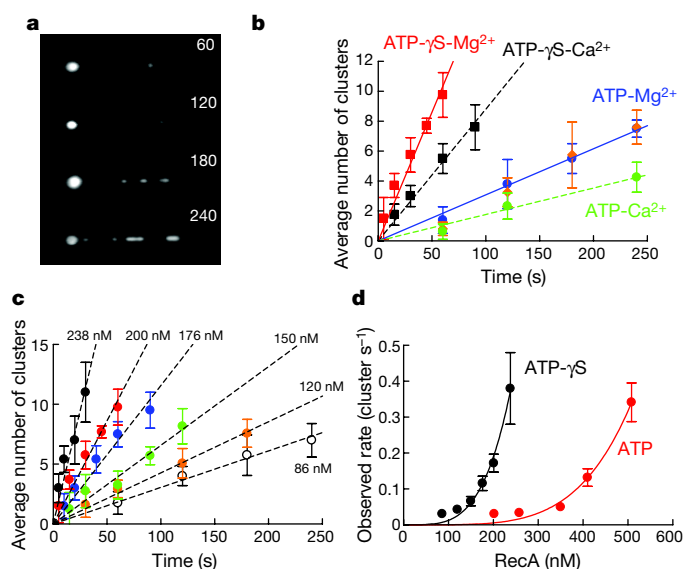


Figure 2 | Nucleation of RecA filaments is independent of ATP hydrolysis but dependent on the type of nucleoside triphosphate and the RecA concentration.

a, Nascent RecA filaments formed by incubation of different DNA molecules with 176 nM RecA, 0.5 mM ATP and 1 mM $\text{Mg}(\text{OAc})_2$ for the indicated times (seconds), and visualized in the observation channel containing 0.5 mM ATP- γ S. **b**, Time dependence of cluster appearance after incubation in the RecA channel with 176 nM RecA and: (red squares) 0.5 mM ATP- γ S, 1 mM $\text{Mg}(\text{OAc})_2$; (black squares) 0.5 mM ATP- γ S, 1 mM CaCl_2 ; (blue circles) 0.5 mM ATP, 1 mM $\text{Mg}(\text{OAc})_2$; (green circles) 0.5 mM ATP, 1 mM CaCl_2 ; and (orange diamonds) 2 mM ATP, 2.5 mM $\text{Mg}(\text{OAc})_2$. Each time point is the average of 4–12 individual molecules. **c**, Time dependence of cluster appearance in 0.5 mM ATP- γ S and 1 mM $\text{Mg}(\text{OAc})_2$ at different RecA concentrations. **d**, Rate of cluster formation as a function of RecA concentration for either ATP- γ S (black circles) or ATP (red circles); the standard deviation for some points is smaller than the symbol size. Error bars in **b–d** indicate standard deviation.

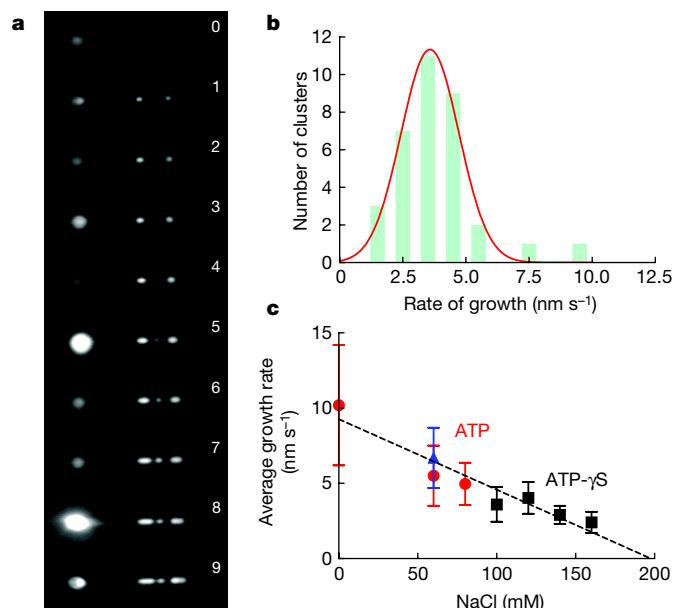


Figure 3 | Growth of individual RecA filaments is independent of nucleoside triphosphate. **a**, One DNA molecule after successive incubations with 170 nM RecA, 0.5 mM ATP-γS, 1 mM Mg(OAc)₂ and 100 mM NaCl. Time in minutes is shown. **b**, Distribution of growth rates. **c**, Salt dependence for the average rate of filament growth observed for 170 nM RecA and: (black squares) 0.5 mM ATP-γS, 1 mM Mg(OAc)₂; (red circles) 0.5 mM ATP, 1 mM Mg(OAc)₂ and stabilized with 5 mM ATP and 10 mM CaCl₂; and (blue triangle) 0.5 mM ATP, 1 mM Mg(OAc)₂ and stabilized with 0.5 mM ATP-γS (see the text). Error bars indicate standard deviation.

and $2.7 \pm 0.2 \text{ nm s}^{-1}$. The distribution of growth rates from individual clusters (Fig. 3b) is $3.6 \pm 1.1 \text{ nm s}^{-1}$ ($n_{\text{clusters}} = 34$) at 100 mM NaCl. When we consider a stoichiometry of 3 bp per RecA monomer and a correction factor of 1.6 due to extension of dsDNA by RecA, the data suggest that filament growth occurs at an apparent rate of ~ 2 RecA monomers s^{-1} under these conditions. Our single-filament rates are comparable to model-derived estimates from single-DNA extension studies when those are normalized for the number of nuclei on a DNA molecule^{4,17,21} (Supplementary Fig. 11).

With ATP, RecA nucleation occurred at a ~ 5 -fold slower rate than with ATP-γS. Consequently, the optimal conditions for monitoring growth from individual ATP clusters are at different salt concentration regimes. However, the salt dependence of growth rates (Fig. 3c) permits a comparison of growth from individual clusters, with either ATP-γS or ATP. The rates range from 3 to 10 nm s^{-1} (~ 2 – 7 monomers s^{-1}) and follow a linear dependence on NaCl concentration. The collinearity of the ATP and ATP-γS rates indicates that the nature of the nucleotide bound to RecA protein has no effect on the elongation phase. Thus, ATP hydrolysis is not involved in filament growth.

When ATP–RecA clusters are stabilized in the observation channel by Ca^{2+} (Supplementary Fig. 12), the subsequent addition of ATP–RecA occurs on nascent filaments that are homogeneous with respect to the bound nucleotide. Alternatively, when the RecA clusters are stabilized by ATP-γS, addition of ATP–RecA to the filament occurs on clusters that are heterogeneous with respect to the bound nucleotide (ATP-γS) (Fig. 3c, blue triangle). The data show that the growth rate is the same whether the RecA clusters are stabilized by Ca^{2+} -ATP or ATP-γS. Thus, this heterogeneous extension experiment shows that both ATP and ATP-γS induce RecA nucleoprotein conformations that are indistinguishable with regard to their capacity to add new RecA molecules.

Assembly of the RecA filament on ssDNA occurs with a net $5' \rightarrow 3'$ polarity²². However, the rates of addition at $5'$ - versus $3'$ -ends are undefined. When a $5'$ or $3'$ ssDNA tail is attached to dsDNA,

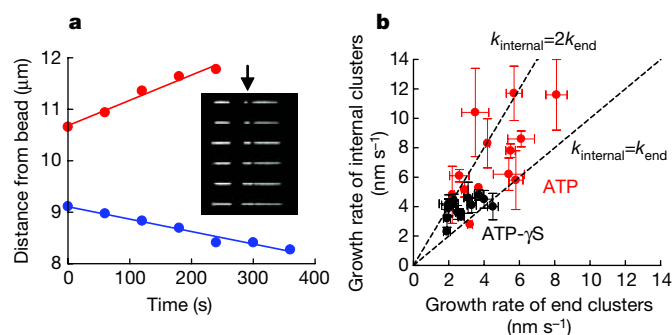


Figure 4 | RecA filaments can grow bidirectionally in the presence of ATP. **a**, Red and blue symbols represent the growing right and left edges of the clusters, respectively; lines are linear fits. The arrow in **a** points to the cluster measured. **b**, Growth rate for RecA filaments that nucleate at an internal DNA location versus the growth rate for filaments that nucleate at the ssDNA end on the same DNA molecule, for all of the salt concentrations examined, for either ATP (red circles) or ATP-γS (black circles). Error bars indicate standard deviation.

polymerization onto the dsDNA of $3'$ -tailed dsDNA is slower than for $5'$ -tailed dsDNA, suggesting that growth on duplex DNA follows the same net $5' \rightarrow 3'$ polarity^{23,24}. Notably, the very fact that RecA can extend in the $3' \rightarrow 5'$ direction suggests that growth can occur at both ends of the filament^{23,24}, albeit at different rates. To address this issue, we analysed molecules where cluster growth was not complicated by the presence of additional clusters growing at neighbouring positions. Of the ~ 140 clusters analysed, only 21 could be unequivocally examined for bidirectionality. A cluster that formed with ATP in the middle of the DNA molecule grew with a strong unidirectional bias (Supplementary Fig. 13); however, we found ATP–nucleoprotein filaments that clearly grew in both directions (Fig. 4a and Supplementary Fig. 13). Although the sample size of these completely isolated clusters is too small to provide meaningful statistics on net polarity of assembly, 19% of the RecA filaments nonetheless assembled bidirectionally with ATP. Although we believe that visualization of bidirectional growth reflects RecA addition to each cluster end, we could not completely exclude the possibility that bidirectional growth originated from clusters bound in opposite orientations but too close to distinguish experimentally. To address this issue further, we compared the rate of growth of clusters that formed with either ATP or ATP-γS at the DNA end and which could only grow from one end in one direction (towards the bead), with clusters that formed internally on the same DNA and, hence, could potentially grow from both ends. We found that the growth rate for single-ended filaments is always lower than or equal to the growth rate for most of the double-ended filaments (Fig. 4b). The higher growth rate for clusters that can extend in both directions provides a kinetic argument for bidirectional filament growth^{23,24}.

Our single-filament observations provide visible evidence that RecA stochastically forms multiple nuclei, which are subsequently extended in the growth phase. ATP hydrolysis is not required for either nucleation or growth, corroborating the proposition that only ATP binding is needed for filament assembly, whereas ATP hydrolysis is needed to facilitate dissociation of RecA either from ssDNA to permit contiguous filament formation or from the heteroduplex DNA product upon completion of DNA strand exchange²⁵. Our analysis of the nucleation and growth phases reveals important differences in these processes. Although neither involves ATP hydrolysis, nucleation is affected by the chemical nature of the nucleotide cofactor, whereas growth is insensitive to the nature of the nucleoside triphosphate bound to either the filament or to the free RecA. Because the affinity of RecA for DNA is greater as a complex with ATP-γS than with ATP¹⁸, this result also supports the idea that nucleation occurs on the DNA molecule. Also, the nucleation rate is profoundly sensitive to the free RecA concentration, suggesting that

4–5 RecA molecules are required to generate a stable cluster on DNA that is extended in the growth phase. The filament is seen to grow at $3\text{--}10\text{ nm s}^{-1}$. Although our results cannot determine whether growth occurs by addition of RecA monomers, if we assume the simplest case of growth by monomer addition, then growth occurs at rates of $\sim 2\text{--}7\text{ monomers s}^{-1}$. This growth can proceed via addition of RecA to both ends of the filament. Ensemble studies suggest that, despite a net growth in the $5'\rightarrow 3'$ direction, growth in the opposite direction is possible^{23,24}. Given that RecA makes contact with only one of the strands of the duplex^{2,23,26}, our findings imply that RecA polymerizes predominantly in the direction imposed by the polarity of the bound ssDNA, with growth in the opposite direction occurring but at a slower rate.

Given the sensitivity of the nucleation phase to solution variables, it is most likely that RecA function is controlled *in vivo* at the nucleation step of filament assembly. This conclusion is in accord with extensive studies that have shown that proteins such as ssDNA-binding protein restrict nucleation²⁷, whereas mediator proteins such as RecBCD and RecFOR promote nucleation^{28,29}. These studies established that mutant RecA proteins, which suppress loss of *recF* function, possess enhanced rates of filament nucleation²⁷. By controlling nucleation, the assembly of RecA on 'normal' cellular ssDNA, such as the lagging strand gaps formed during DNA replication, would be restricted; hence, chromosomal crossovers would be limited. Finally, given the similarity between bacterial RecA and its eukaryotic counterpart, Rad51 (ref. 8), we expect that our approach, results and conclusions regarding the control of filament assembly by nucleation³⁰ apply to mechanistically related processes such as the control of human RAD51 assembly by BRCA2 and other mediators^{9–11}.

METHODS

Visualization of RecA nucleoprotein filaments. The experimental protocol was similar to that reported previously¹⁴. A three-channel flow cell 4.5 mm wide and 70 μm deep generated three 1.5-mm laminar flow paths. Bacteriophage λ DNA, ligated to a 3'-biotinylated oligonucleotide complementary to *cosL*, was attached to a streptavidin-coated 1- μm polystyrene bead (Bangs Laboratories). The bead-DNA complex in 20 mM Tris-OAc (pH 8.2), 20% sucrose and 30 mM dithiothreitol (DTT), containing 10–20 nM YOYO-1, was trapped at a linear flow rate of $100\text{--}130\text{ }\mu\text{m s}^{-1}$ at 25°C for nucleation experiments, or 37°C for growth experiments. The trapped bead- λ DNA complex was then moved to the second, observation, channel containing 20 mM Tris-OAc (pH 8.2), 20% sucrose, 30 mM DTT, 0.2 mg ml⁻¹ catalase, 0.4 mg ml⁻¹ glucose oxidase, 4.5 mg ml⁻¹ glucose, 5 mM Mg(OAc)₂, and the indicated concentration of NaCl, ATP or CaCl₂; for experiments with CaCl₂ in the observation channel, the scavengers and Mg(OAc)₂ were omitted. The YOYO-1 dye allowed measurement of DNA length and verification that a single DNA molecule was captured. In 5 mM Mg(OAc)₂, YOYO-1 dissociated from the DNA after several minutes. The presence of YOYO-1 in this protocol had no effect on RecA nucleoprotein filament properties, because its omission did not change nucleation frequency. The bead- λ DNA complex was then moved to the third channel (RecA channel) containing fluorescent RecA (see Supplementary Information) in 20 mM MES (pH 6.2), 20% sucrose, 30 mM DTT, plus the indicated concentration of NaCl, Mg(OAc)₂, CaCl₂, ATP- γ S or ATP. Finally, after the indicated incubation time in the third channel, the bead-DNA-RecA complex was moved back into the observation channel. Illumination was with a mercury lamp and the fluorescence was observed using an FITC filter (Chroma Technology Corp., number 41001). Analysis is described in Supplementary Information.

Received 25 May; accepted 4 September 2006.

Published online 20 September 2006.

- Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehauer, W. M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
- Kowalczykowski, S. C., Clow, J. & Krupp, R. A. Properties of the duplex DNA-dependent ATPase activity of *Escherichia coli* recA protein and its role in branch migration. *Proc. Natl Acad. Sci. USA* **84**, 3127–3131 (1987).
- Pugh, B. F. & Cox, M. M. Stable binding of recA protein to duplex DNA. Unraveling a paradox. *J. Biol. Chem.* **262**, 1326–1336 (1987).
- Shivashankar, G. V., Feingold, M., Krichevsky, O. & Libchaber, A. RecA polymerization on double-stranded DNA by using single-molecule manipulation: the role of ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **96**, 7916–7921 (1999).

- Hegner, M., Smith, S. B. & Bustamante, C. Polymerization and mechanical properties of single RecA–DNA filaments. *Proc. Natl Acad. Sci. USA* **96**, 10109–10114 (1999).
- Kowalczykowski, S. C. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 539–575 (1991).
- Cazenave, C., Toulme, J. J. & Helene, C. Binding of RecA protein to single-stranded nucleic acids: spectroscopic studies using fluorescent polynucleotides. *EMBO J.* **2**, 2247–2251 (1983).
- Kowalczykowski, S. C. & Eggleston, A. K. Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* **63**, 991–1043 (1994).
- Kowalczykowski, S. C. Cancer: catalyst of a catalyst. *Nature* **433**, 591–592 (2005).
- Sung, P., Krejci, L., Van Komen, S. & Sehorn, M. G. Rad51 recombinase and recombination mediators. *J. Biol. Chem.* **278**, 42729–42732 (2003).
- Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA–ssDNA junction. *Nature* **433**, 653–657 (2005).
- Zaitsev, E. N. & Kowalczykowski, S. C. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* **14**, 740–749 (2000).
- Kim, J. I. & Cox, M. M. The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways. *Proc. Natl Acad. Sci. USA* **99**, 7917–7921 (2002).
- Bianco, P. R. et al. Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374–378 (2001).
- Egelman, E. H. & Stasiak, A. Structure of helical RecA–DNA complexes. Complexes formed in the presence of ATP- γ S or ATP. *J. Mol. Biol.* **191**, 677–697 (1986).
- Sattin, B. D. & Goh, M. C. Direct observation of the assembly of RecA/DNA complexes by atomic force microscopy. *Biophys. J.* **87**, 3430–3436 (2004).
- van der Heijden, T. et al. Torque-limited RecA polymerization on dsDNA. *Nucleic Acids Res.* **33**, 2099–2105 (2005).
- Menetski, J. P., Varghese, A. & Kowalczykowski, S. C. Properties of the high-affinity single-stranded DNA binding state of the *Escherichia coli* recA protein. *Biochemistry* **27**, 1205–1212 (1988).
- Zaitsev, E. N. & Kowalczykowski, S. C. Binding of double-stranded DNA by *Escherichia coli* RecA protein monitored by a fluorescent dye displacement assay. *Nucleic Acids Res.* **26**, 650–654 (1998).
- Kowalczykowski, S. & Steinhardt, J. Kinetics of hemoglobin S gelation followed by continuously sensitive low-shear viscosity. *J. Mol. Biol.* **115**, 201–213 (1977).
- Fulconis, R. et al. Twisting and untwisting a single DNA molecule covered by RecA protein. *Biophys. J.* **87**, 2552–2563 (2004).
- Register, J. C. III & Griffith, J. The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. *J. Biol. Chem.* **260**, 12308–12312 (1985).
- Shaner, S. L., Flory, J. & Radding, C. M. The distribution of *Escherichia coli* recA protein bound to duplex DNA with single-stranded ends. *J. Biol. Chem.* **262**, 9220–9230 (1987).
- Shaner, S. L. & Radding, C. M. Translocation of *Escherichia coli* recA protein from a single-stranded tail to contiguous duplex DNA. *J. Biol. Chem.* **262**, 9211–9219 (1987).
- Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **87**, 21–25 (1990).
- Chow, S. A., Honigberg, S. M. & Radding, C. M. DNase protection by recA protein during strand exchange. Asymmetric protection of the Holliday structure. *J. Biol. Chem.* **263**, 3335–3347 (1988).
- Kowalczykowski, S. C. Biochemical and biological function of *Escherichia coli* RecA protein: behavior of mutant RecA proteins. *Biochimie* **73**, 289–304 (1991).
- Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ -regulated manner. *Cell* **90**, 77–86 (1997).
- Morimatsu, K. & Kowalczykowski, S. C. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol. Cell* **11**, 1337–1347 (2003).
- Spies, M. & Kowalczykowski, S. C. The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. *Mol. Cell* **21**, 573–580 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Spies, J. Siino and A. Forget for suggestions and discussions, and members of the Kowalczykowski laboratory for comments. This work was supported by NIH grants to S.C.K. and R.J.B. R.G. was supported partially by a Fellowship from the Jeanne B. Kempner Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C.K. (skowalczykowski@ucdavis.edu).

naturejobs

Seeing fellow researchers receive awards for their endeavours offers an ideal time to study success and learn valuable lessons about how to guide your scientific career. The two medals given out in the past month by the European Molecular Biology Organization (EMBO), based in Heidelberg, Germany, are no exception.

The organization's 2006 Award for Communication in the Life Sciences went to developmental geneticist Armand Marie Leroi of Imperial College London. Leroi's success had its roots in his desire to spice up his lectures. When he discussed the connection between mutated fruitfly genes and disabilities in humans, his students "sat up and took notice — especially after snoozing through the 50th fly gene", Leroi says. Inspired, Leroi sought out other connections between genetic mutations and human development, and brought them together in a book called *Mutants: On the Form, Varieties and Errors of the Human Body*. A television producer saw a copy of the book's manuscript and approached Leroi to make a programme, which was broadcast under the title *Human Mutants*. Although Leroi says he finds public communication of science "seductive", he tries to stay "grounded" in his research on the nematode worm *Caenorhabditis elegans*. "Ultimately communicating science isn't more satisfying than doing science," he says.

Frank Uhlmann, winner of the 2006 EMBO Gold Medal, attributes much of his success to working in a number of different locations. Now based at Cancer Research UK in London, Uhlmann moved round six labs in Munich while doing his diploma in Germany. He then spent time at the Memorial Sloan-Kettering Cancer Center in New York and Vienna's Research Institute of Molecular Pathology before heading to Britain. "It's always good to go to different places and see different people's approach to research," Uhlmann says. "Learn as much as you can as early as you can."

The career paths of these two winners show how gaining different skills and perspectives can broaden your career. And they also serve as a reminder that success emerges from a passion to do the best science you can.

Paul Smaglik, Naturejobs editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

MOVERS

Luis Serrano, director, Systems Biology Research Unit, Barcelona Biomedical Research Park, Barcelona, Spain



2001-06: Director, Structural and Computational Biology Programme, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
1992-2001: Group leader, EMBL, Heidelberg, Germany

Luis Serrano's self-reliant approach in academia has served him well. Rather than simply attend lectures to pass the end-of-year exam, the biology major at the Complutense University of Madrid created and undertook an independent research proposal — a hands-on exercise that convinced him research was his calling.

In the early 1980s, Serrano pursued graduate work at the Autonomous University in Madrid, focusing on protein microtubules. It was a wise decision given that protein structure and folding were becoming hot topics. There he learned an important lesson for advancing an academic career: construct experiments with a view towards making a small but valuable stepwise contribution, rather than attempt to solve a bigger, less manageable problem in one attempt. Through his PhD work in cell biology and protein biochemistry, and postdoc work on protein folding at both the Autonomous University and the University of Cambridge, UK, Serrano has been at the leading edge of protein biology research. He ultimately chose to go to EMBL in Heidelberg, Germany, for his first senior research position. "I had the best environment to become fully independent, with all the resources I ever wanted," he says.

Serrano has also served as the scientific consultant to four different European companies, advising them at their start-up phase regarding the feasibility of an idea. But the corporate world doesn't hold the same allure for him as academia. "What could a company offer me that I don't have at EMBL?" he asks, noting that a company would dictate the path of his research. Given his independent streak, it's not surprising that Serrano advises young scientists not to blindly follow their supervisor, but to challenge themselves to come up with an original idea. "Become a super researcher, not a super technician," he says.

Now Serrano will direct a new systems-biology unit as part of a joint venture between EMBL and the Centre for Genomic Regulation (CRG) in Barcelona. It will be funded in part by the Spanish Ministry for Education and Science. "The partnership with the CRG will lead to greater interaction between EMBL and the Spanish scientific community, which is in a tremendously dynamic phase of its development," says Iain Mattaj, director-general of EMBL. And as Spain continues to increase spending on new research efforts as part of its National Research and Development Plan, Serrano hopes to use the 'EMBL model' to create a training ground for talented young scientists to become systems biologists. ■
Virginia Gewin

880

RECRUITERS & ACADEMIA

Training peer reviewers

The peer-review process provides an opportunity for clinicians and scientists to train their fellows and postgraduate students as part of a one-on-one journal-club exercise. Supervisors can assess students' critical thinking and writing skills early in their careers. Consider these ten steps to teach the art of reviewing.

1) Explain how to proceed with a peer review, stressing the importance of respecting confidentiality and the rules of the journal or grant body.

2) Establish a submission deadline. The review process should take up to three hours. Set a time for final discussion before the deadline, allowing both student and supervisor time to read and review the work.

3) Adopt the right mindset. Don't accept the manuscript 'as is', and don't be hypercritical. The student might imagine that it's their own manuscript about to be rejected, and that they have the opportunity to improve it.

4) Read the manuscript once, taking rudimentary notes.

5) Follow the journal-specific author and reviewer instructions. Note those pertaining to the category of the manuscript, the word length, abstract structure and the format for references, figures and tables.

6) Verify each citation with PubMed or similar — which also gives

you a chance to read generally on the topic. Highlighting minor citation errors shows that the paper has been read thoroughly. Ask the following: does the submitted work complement or duplicate the authors' and others' previous work?

7) Re-read the manuscript armed with the specific author instructions and good background knowledge.

8) Write the review envisaging that the editor is too busy to read the manuscript in-depth. Summarize the key features in a paragraph, stating the topic of the paper, what was performed, the key conclusions drawn, why this is important, and why this is a novel contribution. Strengths or problems with the manuscript or methodology should then be detailed.

9) Write all comments as if they will be seen by the authors. Although most reviewers are anonymous, caution students that a disgruntled author could recognize one's spelling variances, grammar or clichés.

10) Submit the review, telling the editor the review was written with a student and that you agree with their assessment. Show the student the submitted review as well as the editor's and other reviewers' comments. ■

David A. Mackey is an associate professor of ophthalmology at the University of Melbourne, Australia.

GRADUATE JOURNAL

Computer cold turkey

This week, I suffered one of those losses we all dread. Without any warning, my three-and-a-half-year-old laptop hummed, blue-screened and collapsed. Despite frantic resurrection attempts, it did not recover. Within hours, IT pronounced it dead.

Naturally, such a loss provokes strong emotions. Some of these I was prepared for: the heart-wrenching anxiety of checking my back-ups, and the frustration of hours drudging through the latest computer lingo to find a suitable replacement. But what surprised me was how addicted I had become to the ability to log on any time, any place. Without the freedom to work in cafes and e-mail from home, I felt unproductive and out of touch.

But after the worst withdrawal symptoms had passed, I wondered whether full technological access comes at a cost. An omnipresent tool can make you constantly feel guilty for not working. Before its demise, my laptop would fill my home with whispers: "Come on, your thesis deadline is just three months away! Write, write, write!" Glumly, I'd obey, only to stare torpidly at the screen through the night instead of sleeping. Now, deprived of the 'freedom' of working at home, I'm pushed to make the most of my desktop hours at the office. And for the first time in ages, without my laptop's accusatory glare, free time actually feels free. ■

Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences at the University of Helsinki, Finland.



Faculty Position in Liver Research

The Division Gastroenterology, Hepatology and Nutrition and the Liver Research Group of Cincinnati Children's Hospital Medical Center and Research Foundation invite applications for professorship, at either junior or senior level, for research in liver immunology or liver regeneration.

We seek a candidate with M.D. or Ph.D. degree and outstanding prospects for research, who will complement existing strengths of the Liver Research Group in the fields of developmental immunology of the liver, auto-immune liver disease, and liver regeneration/repair. A successful applicant must be committed to develop and implement independent, externally funded research programs, mentor students, and facilitate interdisciplinary research. The successful applicant will be a member of a highly collaborative group of investigators and will have access to outstanding core resources at the NIH-funded Digestive Health Center and Cincinnati Children's Research Foundation.

Cincinnati Children's Hospital Medical Center and Research Foundation are internationally recognized as one of the nation's top pediatric care and research institutions. The Research Foundation ranks second nationally in NIH funding to full-service Children's Hospitals. Cincinnati is a friendly, pleasant, affordable city with a great quality of life, including many musical and theatrical programs, professional sports, and nearby recreational opportunities.

Academic rank will be determined by candidate's credentials and laboratory experience.

Applicants should submit curriculum vitae and a statement of research interests to:

Dr. Jorge Bezerra, Director of Research, Division of Gastroenterology, Hepatology and Nutrition, CCHMC, 3333 Burnet Ave, Cincinnati, OH 45229-3039, or email to Jorge.bezerra@cchmc.org.

Children's Hospital Medical Center is an Affirmative Action/Equal Opportunity Institution.

Women and minorities are encouraged to apply.

NW90200R



Faculty Positions in Host-Pathogen Interactions

The Ohio State University

Center for Microbial Interface Biology and Department of Microbiology

The Center for Microbial Interface Biology (CMIB) and Department of Microbiology at The Ohio State University invite applications for two tenure-track faculty positions: one Assistant Professor and one Associate Professor. The successful candidates will have a PhD or MD, postdoctoral experience, documented evidence of high quality research, and a commitment to teaching and research at a major research university. They will be expected to establish extramurally funded (Assistant Professor)/have extramurally funded (Associate Professor) independent research programs focused on microbial pathogenesis, the host response to microbes, or experimental therapeutics and vaccines. Expansion by recruitment of additional faculty in these areas will continue over the next several years. The individuals hired will have joint faculty appointments in the Division of Infectious Diseases in the College of Medicine and in the Department of Microbiology in the College of Biological Sciences, as well as membership in the CMIB in the Health Sciences. Very competitive salaries, start-up packages, and excellent biosafety level III, DNA, RNA and protein core facilities are available. Applicants should submit a letter of interest, C.V., description of research plans and the names of at least three potential referees electronically to cmib.search@osumc.edu (preferred) or by mail to: Chair, Search Committee, Center for Microbial Interface Biology, 410 W. 10th Avenue, N-1147 Doan Hall, Columbus, OH 43210. Additional information describing the CMIB and Department of Microbiology can be found at <http://cmib.osu.edu> and <http://www.osumicrobiology.org/>. To ensure full consideration, applications should be received by December 1, 2006, but applications will be accepted until the position is filled. To build a diverse workforce Ohio State University encourages applications from individuals with disabilities, minorities, veterans, and women. Flexible work options are available. EEO/AA employer.

NW90010R



Developmental Mechanisms, Organogenesis, and Stem Cell Research

Faculty positions at Assistant, Associate, and Full Professor levels

Cincinnati Children's Hospital Medical Center, one of the world's top ranking pediatric research centers, is making a major investment in basic and translational research in developmental sciences. The Division of Developmental Biology, under the Directorship of Chris Wylie, will recruit up to twenty additional tenure and tenure-track faculty, to form interdisciplinary groups of basic scientists and clinicians who will study the fundamental mechanisms of development and their application to congenital and acquired childhood disease.

Faculty will occupy outstanding research space in a new 400,000 sq.ft. building, which will be ready in Fall 2007.

Faculty at CCHMC have access to high-quality graduate programs, an MD/PhD program, and training programs for postdoctoral fellows, clinical residents and fellows.

Applications are invited for the first of these positions in the following areas of research:

Fundamental Mechanisms of Development: Candidates working on the molecular and cellular mechanisms of development, using any model animal system, are encouraged to apply.

Organogenesis: Candidates working on the development of the following organ systems are particularly encouraged to apply. Several appointments will be made in each of these areas. Appointment at a senior level could include leadership opportunities in further recruitments.

- Skeletal System
 - Kidney and Urinary Tract
 - Skin
 - Vascular System
-

Stem Cells: Candidates working on any aspect of stem cell research, using any animal model, are encouraged to apply. Up to eight appointments will be made in this area, including a leadership position.

Protein Sciences: We welcome applicants in the general areas of biochemistry, chemical biology, protein engineering/evolution/design, and structural biology who share an interest in studying the molecular basis, diagnosis and treatment of pediatric disorders.

Interested candidates should hold a PhD, MD, or MD/PhD. Please send curriculum vitae, two page research statement, and contact information for three people who will provide letters of recommendation to: DOSpositions@cchmc.org

Candidates should indicate in their cover letters the position(s) in which they are particularly interested.

The Cincinnati Children's Hospital Medical Center, and the University of Cincinnati are Affirmative Action/Equal Opportunity Employers. Qualified women and minority candidates are especially encouraged to apply.

Please visit our web site at: www.cincinnatichildrens.org/research/div/dev-biology

NW90009R



Bowling Green State University

Research Initiative in Photochemical Sciences/Biosciences

The Center for Photochemical Sciences and the Departments of Biological Sciences and Chemistry at Bowling Green State University wish to expand their Research Initiative, interfacing Biology, Chemistry and Photochemical Sciences. Applications are invited for one Assistant Professor tenure-track appointment. The successful candidate will develop and apply photochemical, biophysical, and molecular biology approaches to study the structure, function, dynamics, and interactions of biomolecules and cellular events. Such techniques can include, but are not restricted to, single molecule methods or imaging techniques.

The Department of Chemistry provides a stimulating research environment conducive to faculty success (see <http://www.bgsu.edu/departments/chem/>). Its facilities include the Ohio Laboratory for Kinetic Spectrometry which maintains modern transient spectrometry facilities operating within the femtosecond and longer time domains, and the Wright Photosciences Laboratory which provides unique opportunities for interactions with industrial photoscientists. Dr. H. Peter Lu, the newly-appointed Ohio Eminent Scholar in the Photochemical Sciences, is establishing a Laboratory for Single Molecule Spectrometry, and Dr. Massimo Olivucci, a senior appointee, is initiating a laboratory for computational photochemistry.

The appointment will be made within the Department of Biological Sciences, a PhD-granting department having active research programs in diverse areas of biology (see <http://www.bgsu.edu/departments/biology/>). The successful candidate will have a demonstrated record of productivity in the above mentioned areas and will be expected to develop a strong, extramurally funded research program in his/her specialization. Interested candidates should send a copy of their curriculum vitae, statements of research and teaching interests, representative publications, and three letters of recommendation to the address below.

Applications should be sent to:
Prof. Michael A. J. Rodgers,
Ohio Eminent Scholar and Chair,
Research Initiative Search Committee,
c/o The Department of Biological Sciences,
Bowling Green State University,
Bowling Green, OH 43403.

Review of applications will begin November 15, 2006 and will continue until the position is filled.

BGSU is an Affirmative Action/Equal Employment employer and encourages applications from women, minorities, veterans and individuals with disabilities.

NW90012R



Postdoctoral Fellow

(26UC3422)

The University of Cincinnati

***College of Medicine –
Genome Science***

is accepting applications for a postdoctoral fellow to perform research in molecular biology, biochemistry and cell culture in animal models. Minimum qualifications include a PhD with experience in molecular biology and with zebra fish. Pay is commensurate with experience.

**For additional information and to apply for position #26UC3422, please see
www.jobsatuc.com**

UC is an affirmative action/equal opportunity employer.

NW89882R



Career articles from Naturejobs

- **Naturejobs Prospect:** quick takes on career implications of current events
- **Naturejobs Special Report:** examinations of jobs issues on both sides of the bench
- **Naturejobs Careers & Recruitment:** discipline-by-discipline exploration of opportunities
- **Naturejobs Regions:** tours of scientific hubs
- **Naturejobs Movers:** traffic reports that follow high-profile scientific globe-trotters and sector-hoppers
- **Naturejobs Postdocs & Students:** guides to taking a step toward a permanent position

www.naturejobs.com

naturejobs
making science work

nature publishing group **npg**



Department of Health and Human Services
National Institutes of Health
National Institute of Allergy and Infectious Diseases



THE NATIONAL INSTITUTES OF HEALTH
 @NIH
 OPPORTUNITIES

With nation-wide responsibility for improving the health and well being of all Americans, the Department of Health and Human Services oversees the biomedical research programs of the National Institutes of Health (NIH) and those of NIH's research Institutes.

The National Institute of Allergy and Infectious Diseases (NIAID), a major research component of the NIH and the Department of Health and Human Services, is recruiting for a Tenure/Tenure Track position in the Laboratory of Host Defenses (LHD). The LHD studies immune functions essential for host defense against infection (inherited immune deficiencies) and those required for immune homeostasis (autoimmunity associated with excessive inflammation). The LHD seeks an M.D. or M.D., Ph.D. physician scientist to develop an independent translational research program related to the genetic basis, pathophysiology, diagnosis and treatment of autoimmune diseases associated with excessive inflammation. An emphasis on clinical aspects of innate immunity including phagocytic cells, natural killer cells, dendritic cells and other antigen presenting cells, toll-like receptors or other pattern recognition receptors in its interface with acquired immunity is desirable. The applicant should have a strong track record of basic research of the genetic basis of disease and alterations in signaling pathways responsible for immune dysregulation. The applicant must possess expertise and experience in the design and conduct of diagnostic and therapeutic clinical trials studying and treating autoimmune diseases. Strong clinical credentials in a specialty area relevant to the proposed translational research program (relevant specialties include but are not limited to rheumatology, pulmonary diseases, hematology, immunology or infectious diseases) are required. The program of study proposed by the applicant must include both laboratory components and the conduct of clinical protocols to assess new diagnostic and therapeutic modalities to diagnose and treat autoimmunity associated with excessive inflammation. Applicants particularly suitable for this program are those who have knowledge and experience in the development and clinical application of novel biological agents including chemokines, soluble chemokine receptors, adenosine receptor agonists, monoclonal antibodies, cellular therapies including transplantation or gene therapy to correct the abnormalities in immunity, that achieve immune tolerance or to reduce abnormal inflammation.

The applicant must provide evidence in the submitted materials that the applicant has a current license to practice medicine in one of the states of the United States or must have the all the credentials required by the State of Maryland for licensing to allow the practice of medicine. These credentials must include but are not limited to having a Doctor of Medicine or Doctor of Osteopathy degree from an accredited school in the U.S. or Canada, or a Doctor of Medicine or equivalent degree from a foreign medical school that provided education and medical knowledge substantially equivalent to accredited schools in the U.S. as demonstrated by permanent certification by the Educational Commission for Foreign Medical Graduates (ECFMG).

To be considered for this position, you will need to submit a curriculum vitae, bibliography, three (3) letters of reference, a detailed statement of research interests, and a hardcopy of selected publications to **Thomas A. Fleisher, MD, Chairperson, NIAID Search Committee, c/o Ms. Anissa N. Hunter, DIR Committee Coordinator, Reference Ad #009, 10 Center Drive MSC 1356, Building 10, Rm. 4A26, Bethesda, Maryland 20892-1356 or via e-mail at hunteran@niaid.nih.gov**. Completed applications **MUST** be received by **Thursday, November 15, 2006**. For additional information on this position, and for instructions on submitting your application, please see our website at: **www.niaid.nih.gov**.

NW89846R



THE NATIONAL INSTITUTES OF HEALTH

OPPORTUNITIES @ NIH



HELP US HELP MILLIONS

Department of Health and Human Services (DHHS)
National Institutes of Health (NIH)
National Institute of Allergy and Infectious Diseases (NIAID)

The National Institute of Allergy and Infectious Diseases (NIAID), the second largest institute of the world-renowned National Institutes of Health (NIH), supports and conducts basic and applied research to better understand, treat, and prevent infectious, immunologic, and allergic diseases.

The Dale and Betty Bumpers Vaccine Research Center (VRC), NIAID, NIH is dedicated to translating the latest knowledge of disease pathogenesis and immunology into new vaccine strategies, thereby providing safe and effective means to prevent and control human diseases.



Vaccines for Life™

DALE AND BETTY BUMPERS
VACCINE RESEARCH CENTER
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Department of Health and Human Services

Viral Immunology

The Vaccine Research Center/Biodefense Research Section is **seeking motivated individuals** to study cellular and molecular mechanisms of immune protection from filovirus infection. The primary focus is on using non-human primate models to understand T-cell and humoral anti-viral immunity to Ebola and Marburg virus. Current projects use 12-color flow cytometry and molecular techniques to characterize qualitative and quantitative properties of protective immunity in subjects immunized with gene-based filovirus vaccines (Nature 2000, Nature 2003).

Applicants must have a Ph.D./M.D./D.V.M and a strong background in immunology / virology / microbiology. Experience in cellular or molecular immunology is desirable. Interested applicants should send their cover letter, CV and three letters of reference to Dr. Nancy Sullivan, Chief, Biodefense Research Section, NIH Vaccine Research Center, 40 Convent Drive, Bethesda, MD 20892 or by e-mail (nsullivan@nih.gov) for consideration.



The National Institute of Allergy and Infectious Diseases, a major research component of the NIH and the Department of Health and Human Services, is recruiting a Staff Scientist. The position will be available in the Respiratory Viruses Section of the Laboratory of Infectious Diseases, and scientists with a M.D., D.V.M, or Ph.D. are eligible. The research activity involves (1) examination of the pathogenesis of pandemic and potential pandemic strains of influenza and their evaluation in vitro and in experimental animals; (2) influenza viral genomics, and examination of viral evolution in fitness and host adaptation; and (3) the development of influenza clinical trials in humans. This full-time research position offers a unique opportunity to work on investigations that range from basic molecular biology to clinical research. Staff Scientist applicants should have at least six years of laboratory work experience in molecular and classical virology research; the salary range is \$73,178 - \$165,195. Preference will be given to candidates who have experience working with avian influenza viruses and those with BSL3 experience. Applicants should submit their curriculum vitae, a letter of research interests, and names and addresses of three references to: **Jeffery K. Taubenberger, MD, PhD, Attn: D. Kyle, NIAID, NIH, Bldg 50 Room 6234, MSC 8007, 50 South Drive, Bethesda, MD 20892-8007, FAX: (301) 496-8312, email: dkyle@niaid.nih.gov.** Review of applicants will begin on **October 30, 2006** and continue until a successful candidate is identified.

Postdoctoral, Research, and Clinical Fellowships at the National Institutes of Health

www.training.nih.gov/pdopenings

www.training.nih.gov/clinopenings

Train at the bench, the bedside, or both

Office of Intramural Training and Education
Bethesda, Maryland 20892
800.445.8283



DIRECTOR, GRADUATE PARTNERSHIPS PROGRAM
Office of Intramural Training and Education
National Institutes of Health



The Intramural Research Program of the National Institutes of Health in Bethesda, Maryland is seeking an outstanding individual to serve as Director of the Graduate Partnerships Program (GPP), Office of Intramural Training and Education (OITE). The Director, Graduate Partnerships Program, oversees the organization, implementation, and monitoring of Ph.D.-granting graduate programs associated with the NIH Intramural Research Program and assures that the partnerships are of the highest quality in training, curriculum and student outcomes. Specifically, the GPP is responsible for developing and monitoring policies and procedures and for coordinating institutional support to ensure that graduate student training programs meet the specific criteria of the universities involved in Ph.D.-granting graduate partnerships with the NIH. The incumbent also has a major responsibility for the design and implementation of strategies to facilitate the recruitment of students into the various NIH graduate partnerships. The Director chairs the GPP Executive Committee and oversees an office of ~7 staff. A Ph.D. or equivalent degree plus scientific knowledge and demonstrated expertise in biomedical science as well as experience in running a research group and working with trainees, including graduate and undergraduate students and postdoctoral fellows, is required. A scientific background is also essential for working with NIH and university faculty mentors.

Further details about GPP and specific programs can be found at www.training.nih.gov. The salary range is \$107,521 - \$139,774 and a full package of benefits is included. A detailed vacancy announcement with the mandatory qualifications and application procedures can be obtained on USAJOBS at www.usajobs.gov (announcement number **OD-06-14059**) and NIH Web Site at <http://www.jobs.nih.gov>. Questions on application procedures may be addressed to **Laurie Steinman** on (301) 594-5335. Applications must be received by midnight eastern standard time on **December 28, 2006**.



Tenure Track Investigator
for Symptom Management Laboratory

National Institute of Nursing Research, NIH
 NINR seeks applications from nurse-scientists for a tenure-track or tenure-eligible position in its Symptom Management Laboratory. The successful applicant will be expected to establish an innovative independent research program in symptom management in chronic or acute illness. Individuals with documented expertise in biobehavioral approaches to symptom management research are of particular interest. Candidates must have a Ph.D. as well as advanced training and accomplishment in symptom management research. The NINR Symptom Management Laboratory offers unparalleled opportunities for multidisciplinary collaboration throughout NIH.

For additional information, contact **Renee G. Harris** at harrisrg@mail.nih.gov.

To apply, send your CV, cover letter including a summary of relevant experience, and the names and contact information for 3 recent references to: **Renee G. Harris, National Institute of Nursing Research, NIH, 10 Center Drive, CRC-East, Room 2-1339, Bethesda, MD 20892**. Applications will be accepted until the position is filled.

DEPARTMENT OF HEALTH AND HUMAN SERVICES (DHHS)
NATIONAL INSTITUTES OF HEALTH (NIH)
NATIONAL CANCER INSTITUTE (NCI)



HIV and AIDS Malignancy Branch
Center for Cancer Research

Tenured/Tenure Track Position
Translational Researcher in Viral Oncogenesis

The HIV and AIDS Malignancy Branch (HAMB), NCI, is searching for a tenure-track or tenured investigator in the field of viral oncogenesis. It is anticipated that the investigator will establish an independent research program targeted to the study of the treatment, pathogenesis, and/or prevention of viral-induced tumors, especially those associated with AIDS. The research program should be translational in focus and be able to interface with a strong existing clinical research program in AIDS-related tumors. HAMB is located on the Bethesda campus of the NIH (<http://ccr.cancer.gov/labs/lab.asp?labid=63>). Current areas of research in HAMB focus on Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8)-associated tumors, the molecular biology of human papillomavirus (HPV), and the development of novel therapeutic interventions for HIV infection. Candidates for the position should have an M.D./Ph.D., Ph.D., or M.D. and strong research credentials. Applicants for this position should submit a curriculum vitae including bibliography, a statement of research interests, a two-page outline of the proposed research program, and the names of three references to **Chairman, Search Committee, HAMB, NCI, Attention Jan Huque, Building 10, Rm.10S255, 10 Center Drive, M.S.C. 1868, Bethesda, MD 20892-1868** no later than **December 7, 2006**. You may also e-mail your application to: huquei@mail.nih.gov (Jan Huque, 301-435-4627).

NW89845R

A brief history of death switches

Immortal remains.

David Eagleman

There is no afterlife, but a version of us lives on nonetheless.

At the beginning of the computer era, people died with passwords in their heads and no one could access their files. When access to these files was critical, companies could grind to a halt. That's when programmers invented death switches.

With a death switch, the computer prompts you for your password once a week to make sure you are still alive. When you don't enter your password for some period of time, the computer deduces you are dead, and your passwords are automatically e-mailed to the second-in-command. Individuals began to use death switches to reveal Swiss bank account numbers to their heirs, to get the last word in an argument, and to confess secrets that were unspeakable during a lifetime.

It soon became appreciated that death switches provided a good opportunity to say goodbye electronically. Instead of sending out passwords, people began programming their computers to send e-mails to their friends announcing their own death. "It appears I'm dead now," the e-mails began. "I'll take this as an opportunity to tell you things I've always wanted to express..."

Soon enough, people realized they could program messages to be delivered on dates in the future: "Happy 87th birthday. It's been 22 years since my death. I hope your life is proceeding the way you want it to."

With time, people began to push death switches further. Instead of confessing their death in the e-mails, they pretended they were not dead. Using auto-responder algorithms that cleverly analysed incoming messages, a death switch could generate apologetic excuses to turn down invitations, to send congratulations on a life event, and to claim to be looking forward to a chance to see them again sometime soon.

Today, building a death switch to pretend you are not dead has become an art form. Death switches are programmed to send a fax occasionally, make a transfer between bank accounts, or make an online purchase of the latest novel. The most sophisticated switches reminisce about shared adventures, exchange memories about a good story, swap inside jokes, brag about past feats, summon up lifetimes of experience.

In this way, death switches have estab-

lished themselves as a cosmic joke on mortality. Humans have discovered that they cannot stop Death, but at least they can spit in his drink.

This began as a good-spirited revolution against the grave's silence. The problem for those of us still living, however, is the increasing difficulty in sorting the dead from the living. Computers operate around the clock sending out the social intercourse of the dead: greetings, condolences, invitations, flirtations, excuses, small talk, inside jokes — codes between people who know each other well.

And it is clear now where this society is going. Most people have died off, and I'm one of the few remaining. By the time I die and my own death switch is triggered, there will be nothing left but a sophisticated network of transactions with no one to read them: a society of e-mails zipping back and forth under silent satellites orbiting around a soundless planet.

So an afterlife does not exist for us, *per se*, but instead an afterlife exists for that which exists between us. When an alien civilization eventually bumps into Earth,

it will immediately be able to understand what humans were about, because what will remain is the network of relationships: who loved whom, who competed, who cheated, who laughed together about road trips and holiday dinners. Each person's ties to bosses, brothers, lovers are written in the electronic communiqués. The death switches simulate the society so completely that the entire social network is reconstructable. The planet's memories survive in zeros and ones.

This situation allows us to forever revisit shared jokes, to remedy lost opportunities for a kind word, to recall stories about delightful earthly experiences that can no longer be felt. Memories now live on their own, and no one forgets them or grows tired of telling them. We are quite satisfied with this arrangement, because reminiscing about our glory days of existence is perhaps all that would have happened in an afterlife anyway.

David Eagleman is a neuroscientist who studies how the brain represents time. He also writes fiction. He can be found at www.eaglemanlab.net.

